8-2013

Towards Viable Large Scale Heterogeneous Wireless Networks

Rahul Amin
Clemson University, ramin@clemson.edu

TOWARDS VIABLE LARGE SCALE HETEROGENEOUS WIRELESS NETWORKS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Computer Engineering

---

by
Rahul Amin
August 2013

Accepted by:
Dr. James Martin, Committee Chair
Dr. Harlan B. Russell, Committee Co-Chair
Dr. Daniel Noneaker
Dr. Brian Dean
Dr. Melissa Smith

ABSTRACT

We explore radio resource allocation and management issues related to a large-scale heterogeneous (hetnet) wireless system made up of several Radio Access Technologies (RATs) that collectively provide a unified wireless network to a diverse set of users through co-ordination managed by a centralized Global Resource Controller (GRC). We incorporate 3G cellular technologies HSPA and EVDO, 4G cellular technologies WiMAX and LTE, and WLAN technology Wi-Fi as the RATs in our hetnet wireless system. We assume that the user devices are either multi-modal or have one or more reconfigurable radios which makes it possible for each device to use any available RAT at any given time subject to resource-sharing agreements. For such a hetnet system where resource allocation is coordinated at a global level, characterizing the network performance in terms of various conflicting network efficiency objectives that takes costs associated with a network re-association operation into account largely remains an open problem. Also, all the studies to-date that try to characterize the network performance of a hetnet system do not account for RAT-specific implementation details and the management overhead associated with setting up a centralized control. We study the radio resource allocation problem and the implementation/management overhead issues associated with a hetnet system in two research phases. In the first phase, we develop cost models associated with network re-association in terms of increased power consumption and communication downtime taking into account various user device assumptions. Using these cost models in our problem formulations, the first phase focuses on resource allocation strategies where we use a high-level system modeling approach to study the

achievable performance in terms of conflicting network efficiency measures of spectral efficiency, overall power consumption, and instantaneous and long-term fairness for each user in the hetnet system. Our main result from this phase of study suggests that the gain in spectral efficiency due to multi-access network diversity results in a tremendous increase in overall power consumption due to frequent re-associations required by user devices. We then develop a utility function-based optimization algorithm to characterize and achieve a desired tradeoff in terms of all four network efficiency measures of spectral efficiency, overall power consumption and instantaneous and long-term fairness. We show an increase in a multi-attribute system utility measure of up to 56.7% for our algorithm compared to other widely studied resource allocation algorithms including max-sum rate, proportional fairness, max-min fairness and min power. The second phase of our research study focuses on practical implementation issues including the overhead required to implement a centralized GRC solution in a hetnet system. Through detailed protocol level simulations performed in ns-2, we show an increase in spectral efficiency of up to 99% and an increase in instantaneous fairness of up to 28.5% for two sort-based user device-to-Access Point (AP)/Base Station (BS) association algorithms implemented at the GRC that aim to maximize system spectral efficiency and instantaneous fairness performance metrics respectively compared to a distributed solution where each user makes his/her own association decision. The efficiency increase for each respective attribute again results in a tremendous increase in power consumption of up to 650% and 794% for each respective algorithm implemented at the GRC compared to a distributed solution because of frequent re-associations.

DEDICATION

This thesis is dedicated to my beloved family - my mother, my father and my brother. Their unconditional love and support has helped me climb the steps of success in every phase of my life.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Jim Martin, for all his guidance and motivation throughout my PhD studies because of which the work presented in this dissertation has been possible. I would like to thank him for all the valuable knowledge and experience he has shared with me. I would also like to thank Dr. Harlan B. Russell, Dr. Daniel L. Noneaker, Dr. Brian Dean and Dr. Melissa Smith for serving on my committee. Finally, I would like to thank Dr. Ahmed Eltawil and Amr Hussien from University from California, Irvine, and Dr. Luiz DaSilva and Juan Deaton from Virginia Tech who collaborated with us on several related research topics.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER ONE

INTRODUCTION

Advances in wireless technology as well as in microelectronics and battery technology have helped to connect the world in unprecedented ways. According to recent PEW surveys, 85% of Americans use a cell phone, 43% of them access the Internet with their cellular device, and at least 35% of them own a smartphone [1-2]. It is expected that by 2013 multimodal smartphones will overtake PCs as a means to access the Internet [3]. Ericson's most recent Traffic and Market Data report predicts that global mobile data traffic will grow tenfold by 2016. The number of Wi-Fi hotspots will triple by 2015 [4]. The demand for wireless access goes beyond cellular device access. It is expected that application specific domains such as eHealth, smart grid, intelligent transportation systems, and environmental sensing will lead to potentially very large scale adoption of Machine-to-Machine technology [5]. The Internet community has identified an 'Internet of Things' concept where the internet will have to support a tremendous number of devices, in particular 'machines' that require wireless connectivity [6-7]. In part, due to this exponential growth, the FCC has projected that the nation's wireless operators will face a 275 MHz spectrum deficit by 2014 if no new spectrum is made available for broadband usage; this has motivated federal mandates to add 500 MHz of spectrum [8]. However, studies show that while areas of spectrum are over-utilized, there are areas of spectrum that are under-utilized [9-10]. This under-utilization of spectrum has renewed interest in techniques or paradigms such as co-operative communications, symbiotic networking, cognitive networking, and dynamic spectrum access which attempt to

improve spectral efficiency through co-operation at the radio level [11-28]. While symbiotic networking, cognitive networking, and dynamic spectrum access methods focus on improving efficiency from the bottom up with regards to the OSI stack, heterogeneous wireless networks represent methods of co-operation driven from the top down. This latter heterogeneous wireless networks approach is the focus of our study.

1.1 Terminology

We define the term Autonomous Wireless System (AWS) to mean an independent wireless network that is administered by a single management authority, such as a cellular operator/carrier. We simplify our discussion by assuming that an AWS implies one RAT. Therefore, an organization's 802.11 network or a public Wi-Fi hotspot are both examples of a single AWS. A large LTE-Advanced network that internally uses heterogeneous components such as relays or picocells is also a single AWS. We define a heterogeneous wireless network (referred to as a hetnet) as a wireless system that typically involves more than one RAT and involves more than one AWS. From the user device perspective, we use the term cognitive and reconfigurable device interchangeably. Each cognitive/reconfigurable device is capable of connecting to and supporting data transmission over more than one RAT. Our definition of a cognitive device (or radio) differs slightly from the generally accepted version of a cognitive radio that identifies it as "a radio frequency transceiver designed to intelligently detect whether a particular segment of the radio spectrum is in use, and to jump into (and out of) the temporarily unused spectrum very rapidly, without interfering with the transmission of other authorized users" [21]. In our work, we assume that a single management authority (such

as AP/BS) for each AWS already deals with spectrum usage details and so our cognitive devices are not required to sense and find open spectrum, but rather they need to be capable of switching access modes as required by the decision-making entity.

1.2 Research Motivations and Direction

Due to widespread deployment of wireless access technologies, it is quite common for any geographical location to be covered by more than one wireless network. The number of wireless networks in any given area is expected to grow for at least the following reasons: the trend for open Wi-Fi access will continue; RATs involving open spectrum are likely to become available; and as 4G evolves, the number of legacy systems will grow. So a user device at any given location will have multiple connectivity options; the number of connectivity options per device will be limited by the number of radios equipped on the device and in case of cellular networks, the network usage agreement (subscription) the user has signed up for. While emerging user devices are expected to support a multitude of wireless access methods, the current access methods *require the user to select the active access network* either by purchasing an appropriate handset (and service) or, in the case of multimodal smartphones, by manually selecting the access network. Once the user selects the access network, each network attempts to achieve the best performance within its own network, generally ignoring impacts of co-located wireless networks. Localized resource allocation decisions will usually not lead to optimal resource usage. For example, [29] shows that the 'selfish' approach can result in non Pareto-optimal bandwidth allocation as compared to the case where a centralized entity performs network-wide resource allocation. A non Pareto-optimal allocation is one such

that, there exists another feasible allocation where at least one user gets more bandwidth, and all others get at least the same bandwidth. Significant improvements in efficiency result when the resource management process jointly considers the distribution of resources across network technologies, reaping the benefits of multi-access network diversity. *A fundamental motivation of our research is that enhancing access and use of spectrum requires a combination of cognitive device capabilities AND a component of the resource allocation control that operates at the global level*. Inherent in this problem is ensuring the methods scale and co-exist with standards-based equipment.

We address the very timely issue of how multiple AWSs can co-operate and collectively guide agile or reconfigurable devices to improve the efficiency with which spectrum is used by seamlessly directing these devices to select the most efficient RAT available among a number of possible RATs. To achieve this goal and to focus the research, our work addresses how one can build large-scale unified wireless networks that leverage sophisticated device modalities. We assert two necessary assumptions: 1) each user device is capable of supporting multiple RATs offering varying degrees of service and quality attributes that the network is aware of, and through the use of a centralized GRC, the network can instruct the device to change its access mode as needed to enhance the efficiency by which spectrum is utilized; 2) incentives are in place motivating independent AWSs to co-operate to provide users a network with enhanced coverage and performance, which is better than what could be achieved by any single AWS.

To ground the research to current state of the art, we use several illustrative examples. In the first example, “Wi-Fi offloading” refers to how cellular systems

internetwork with 802.11 Wi-Fi networks. Currently, cellular systems leave the choice of access to the end user; however, cellular operators would prefer their customers to use Wi-Fi whenever available. The benefits from this approach are multi-fold. First, the cellular operator saves the expensive macro-cell capabilities for the truly mobile members of the cell. Second, the performance for low mobility, indoor members of the cell is improved by avoiding indoor penetration issues, thus significantly improving overall network performance. While some commercial carriers (e.g. T-Mobile and others) have already experimented with this approach, maintaining a seamless transition between the networks has proven elusive to-date because the Wi-Fi network is typically out of the operator's control [30]. The second example relates to femto-cell deployments, which have been proposed as a method to increase spectral efficiency by supplementing the macro-cell with an overlay of smaller, co-operative networks [31-32]. A broadband access network is utilized to backhaul the cellular traffic back to the wireless operator. The two examples are similar in spirit as they attempt to improve the connectivity of handheld devices and to offload traffic from the macro-cell. Both examples illustrate a clear direction - independent networks co-operating. Note that the femto-cell example requires advanced interference management as both the femto-cell and macro-cell utilize the same frequency band. Our work is closer to the Wi-Fi offloading example as we assume all RATs in our hetnet system operate on a separate frequency band and hence do not interfere with other co-located RATs.

Another scenario of co-operation allows users of one cellular operator access to another operator's infrastructure through peering agreements. The wired Internet is based

on co-operative agreements between service providers. The economics related to wireless networks is very different. There are FCC guidelines that identify roaming arrangements between wireless operators when infrastructure is not available. However, the rules are easily subverted by wireless operators charging exorbitant roaming rates. A core conjecture motivating our research is that network co-operation between wireless operators would benefit end users and in turn provide new economic opportunities for operators. Throughout our work, we do not show an increase in (monetary) profit for network operators, which usually drive any resource sharing agreements. Rather, we assess the benefits of co-operation by showing an increase in achieved performance in terms of network efficiency measures of spectral efficiency, instantaneous and long-term fairness, and overall power consumption.

1.3 Research Objectives

The overall research related to the implementation of a large-scale hetnet wireless system is divided into two phases. The first phase focuses on the *resource allocation* problem in a hetnet wireless system which concentrates on balancing the conflicting network efficiency objectives of maximizing spectral efficiency, maximizing instantaneous and long-term fairness, and minimizing overall power consumption. While there have been several works that focus on the network selection problem in a heterogeneous wireless network based on multiple performance objectives [29,33-40], the work in [29,33] comes closest in terms of the proposed centralized solution and the network efficiency parameters considered in our work. The work in [29] proposes a generalized proportional fairness resource allocation scheme that obtains an acceptable

tradeoff between throughput and user fairness for a system that makes user device-to-BS association decisions on a global level. In [33], the authors propose a vertical handover decision algorithm implemented at a centralized handoff controller that tries to attain a tradeoff in terms of the conflicting objectives of maximizing collective battery lifetime of user devices and the load balancing criteria of APs/BSs. However, none of these works account for costs associated with a network re-association operation in analyzing their proposed solutions. We define network re-association operation as a process a user device has to undergo to re-establish wireless connectivity when it switches its association from one AP/BS to another AP/BS. The use of network re-association cost models in analyzing achievable performance in terms of conflicting network efficiency measures of spectral efficiency, overall power consumption, and instantaneous and long-term fairness for each user in the hetnet system is the main focus of our work conducted in first phase of research study. Within the first phase of our work, we perform two studies that have different research objectives. For the first study, using a heuristic resource allocation algorithm, we address the following research questions: What impacts do various network topologies and user device assumptions have on achieved network efficiency measures of spectral efficiency and overall power consumption? What impact does the number of re-associations under different network co-operation models have on overall power consumption? How many reconfigurable radios are required per user device to achieve the benefits of network co-operation? For the second study, using a limited set of network topologies and user device assumptions, we use an optimization-based resource allocation approach where we address the following research questions: How can the resource

allocation procedure achieve desired tradeoffs in terms of all four conflicting network efficiency objectives? How do the achieved tradeoffs compare to the network efficiency performance measures achieved by traditional algorithms proposed for hetnet systems?

The second phase of our research study focuses on *practical implementation issues* including the modeling of overhead required to implement a centralized GRC solution in a hetnet system. The IEEE 802.21 framework, which emerged from the Always Best Connected (ABC) concept, has been used as a basis for the control plane required for a hetnet system. The key goal of the ABC concept is to enable user devices to seamlessly switch to the best RAT when multiple RATs are available. Research on ABC and IEEE 802.21 has primarily focused on seamless handover to the best available network to minimize latencies associated with network re-association [41-48]. An analytical model to estimate the amount of overhead in a centralized hierarchical wireless system is studied in [49]. But to the best of our knowledge, no research that attempts to characterize the performance of a hetnet system in terms of network efficiency metrics accounts for RAT-specific implementation details and the management overhead associated with setting up a centralized control. For the second phase of our research study, we focus on using a detailed protocol level simulator, ns-2, to model each RAT and the control plane (message overhead due to centralized control) within the hetnet system using IEEE 802.21's media independent handover function. Using this setup, we address the following research questions: What are the performance gains for a centralized solution in a hetnet system compared to a distributed solution in terms of spectral efficiency and (instantaneous and long-term) fairness? What technical challenges are

associated with the implementation of a centralized hetnet solution? How much increase in power consumption is caused by network re-associations for a centralized solution compared to a distributed solution? How much overhead is caused by the centralized solution relative to the overall system (data) throughput?

1.4 Dissertation Outline

The rest of the dissertation is organized as follows. We present relevant background related to radio resource management architectures/frameworks and resource allocation techniques for hetnets in Chapter 2. We present our system description in Chapter 3 where we describe our system model, network co-operation model, and define our network performance measures. We present the work focused on resource allocation problem considered in the first phase of our research study in Chapter 4. We present the work related to practical implementation issues, such as modeling the overhead required for a centralized GRC solution, conducted in the second phase of our research study in Chapter 5. We conclude our work in Chapter 6.

CHAPTER TWO

BACKGROUND

The creation of a hetnet system managed by a centralized GRC requires identifying two solutions: 1) overhead management 2) resource allocation techniques. The overhead management involves information exchange between the user devices, resource controllers (BS/AP) of each RAT and the GRC. Once the information related to each RAT and user device is available at the GRC, the GRC makes the user device-to-AP/BS association decisions based on the implemented resource allocation procedure. We provide an overview of radio resource management frameworks that establish guidelines to manage overhead in a hetnet system and provide a literature survey on various resource allocation techniques proposed for a hetnet system in this chapter.

2.1 Radio Resource Management Frameworks

At the network level, IEEE and 3GPP standardization bodies have suggested architectures and frameworks to support hybrid or heterogeneous networks [50-52]. A survey of these architectures has been provided in [52]. Recent proposals have been based on the media independent handover function defined by the IEEE 802.21 standard which provides a framework to support seamless mobility through networks based on different radio access technologies without the need to restart the radio connection every time the mobile moves to a new network [50]. Another relevant standard, IEEE P1900.4, defines building blocks for enabling coordinated network-device distributed decision making, which will aid in the optimization of radio resource usage, including spectrum access control, in heterogeneous wireless access networks [51]. Hierarchical resource managers

have been proposed by the Common Radio Resource Management, Joint Radio Resource Management and Multi-access Radio Resource Management schemes studied by the 3GPP group. In these hierarchical schemes, and also in our proposed system, the local resource managers of different wireless technologies interact with a centralized entity to jointly optimize the process of resource allocation. The presented IEEE and 3GPP frameworks have been used as basis for building heterogeneous wireless systems [53-55]. Finally, perhaps the most relevant frameworks are the recent standards being developed by the IETF and 3GPP communities to support 'flow mobility' as a mobile user roams over multiple wireless access systems [56-58].

To fully benefit from the emerging hetnets concept where multiple RATs are managed by a centralized resource coordinator, there is also a need for efficient design-time and run-time reconfigurable platforms. Numerous reconfigurable architectures have been proposed spanning different technologies including application specific instruction set processors (ASIPs), field programmable gate arrays (FPGAs), and digital signal processors (DSPs). Recently, multi-processor systems on chip (MPSoC) architectures have evolved rapidly in the race of high performance embedded computing [59], especially in applications that require a flexible computing structure that can be reconfigured to handle various applications. A common design metric among all platforms is reducing power consumption that restricts both the capabilities of the device and the design choices that are available. Towards that end, numerous techniques have been developed to optimize power consumption at different levels including algorithm, system, architecture, and circuit levels [60-61]. Hence, enough progress has been made at

both the system architecture level and at the user device level to make the implementation of a real hetnet system managed by a centralized controller feasible in the near future.

2.2 Resource Allocation: Scheduling Perspective

An inherent component of a resource allocation method is the objective or the strategy that is used to guide the allocation decisions. We consider the resource allocation objectives of maximizing spectral efficiency (or system throughput), instantaneous and long-term fairness, and overall power consumption. The majority of resource allocation approaches for wireless networks consider trade-offs between throughput maximization and fairness objectives while allocating resources to competing users. There are two generally accepted resource allocation policies for wireless networks: max-min fairness and proportional fairness. Max-min fairness resource allocation procedure, which can be obtained via progressive-filling algorithm, allocates rates to users such that it is not possible to increase the rate of any user in the system without decreasing the rate allocated to any other user who is receiving an already lower or equal rate [62]. This approach tries to maximize fairness among rates allocated to all users in the system, but in doing so sacrifices achievable spectral efficiency. Proportional fairness resource allocation procedure maximizes the sum of the log of rates allocated to each user in the system [63]. This allocation procedure is designed to take advantage of multiuser diversity while maintaining comparable long-term fairness for all users. This procedure is an accepted tradeoff in terms of maximizing throughput and long-term fairness objectives and is implemented in current cellular systems.

Current cellular systems deploy a hierarchy of resource controllers. Each device,

along with its assigned base station, independently tries to optimize the resource allocation process within its own domain, generally ignoring impacts of co-located heterogeneous wireless networks. Localized resource allocation decisions will usually not lead to optimal resource usage. A key motivation for our work is the fact that most current design practices still involve building independent RATs. The work in [29] recognizes that while a significant amount of research has explored the use of proportional fairness for resource allocation in wireless systems, all studies have focused on fairness maintained by a single base station. The authors share our motivations that the association of devices to specific networks must be done at a global level in order to maximize network efficiency.

Moreover, proportional fairness scheduling does not satisfy minimum instantaneous data rate requirements of real-time traffic, as it allocates all the resources in a scheduling interval to the user device with the highest achievable ratio of instantaneous to average data rate [64]. Today's mobile internet applications, such as voice, video, gaming and social networking services, have diverse traffic characteristics and, consequently, have different Quality of Service (QoS) requirements. As a result, a QoS framework is a fundamental component of current 4G and next generation wireless networks. In addition to the best-effort service, 4G standards LTE and WiMAX define various service classes, such as Guaranteed Bit Rate (GBR), Unsolicited Grant Service (UGS) and Real-time Polling Service (rtPS), which have minimum data rate requirements per scheduling interval [65]. So instantaneous fairness metric has to be considered by any resource allocation procedure intended for future hetnet systems.

At the same time, environmental concerns and user device requirements have elevated the importance of energy efficient networks and devices. As wireless operators have learned, a handheld device's battery efficiency is a very visible attribute of an operator's services [66]. Unfortunately, in many situations, methods for improving spectral efficiency directly lead to an increase in power consumption. Recently, due to a renewed interest in 'green communications' and users' increased expectations from mobile device battery life, researchers have started focusing on minimizing overall power consumption subject to fairness constraints and other network-efficiency requirements, such as throughput and delay [67-69]. However, to the best of our knowledge, none of these works have looked at the trade-offs surrounding power consumption, spectral efficiency, and fairness in large-scale heterogeneous wireless networks that involve reconfigurable user devices or that involve different assumptions surrounding the level of co-operation available between underlying independent wireless systems.

2.3 Resource Allocation: Optimization Perspective

Significant effort has gone into the joint optimization of the resource allocation process in a cellular (or WLAN) system constrained by combinations of fairness, spectral efficiency and power requirements [67-72]. More recent effort has gone into the 'network selection' problem which describes the method by which a client device determines when to initiate a vertical handoff and which network should be joined. Network selection algorithms for optimal service delivery over user devices capable of connecting with several RATs can be categorized into several strategies: decision function-based strategies, user-centric strategies, multiple attribute decision-making strategies, and fuzzy

logic and neural networks-based strategies. All these strategies use a set of attributes in the decision making process which are either related to the user or to the service provider. Some of the user-related attributes include achieved throughput by each individual user, battery lifetime of each mobile terminal, and QoS parameters such as packet delay, jitter and loss. Service provider-related attributes include load-balancing, throughput fairness amongst users, incurred cost per transmitted data byte, and overall revenue. The decision function-based strategies use a weighted utility function that incorporates both user-related and service provider-related network selection attributes [33-34]. The user-centric strategies focus on one or more needs of the user to decide on the choice of current access network [35]. Multiple Attribute Decision Making (MADM) deals with the problem of choosing from a set of alternatives that are characterized in terms of their attributes. The most popular classical MADM methods are Simple Additive Weighting (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), and Grey Relational Analysis (GRA). A comparison of these models was established in [36] with bandwidth, delay, jitter, and bit error rate attributes. It showed that SAW and TOPSIS provide similar performance under all traffic classes examined. GRA provides slightly higher bandwidth and lower delay to interactive and background traffic classes. Fuzzy logic and neural network concepts are applied to choose when and to which network to hand-off among different available access networks when a decision problem contains attributes with imprecise information [37-38].

All the strategies for network selection algorithms described in the previous paragraph make use of multiple user-related or network provider-related attributes. The

method to determine the relative importance of each attribute under consideration has significant impact on the solution space and the implementation complexity of the algorithm. Several related works have looked at multiple weight combinations of the attributes based on imprecise user preferences [34,39,40]. Other works have selected the attribute weights based on simulation results by determining the difference in magnitude of each attribute and then assigning each attribute equal importance [33,38]. In the multi-attribute optimization based resource allocation study we perform in the first phase of our research, we reduce the solution space of our algorithm by using responses from network provider interviews and the Analytical Hierarchy Process (AHP) [73] to determine the relative weights of each attribute in our optimization function.

Game theory has also been employed to model the network selection problem. The authors of [74] propose a network selection scheme to accommodate current demand and minimize handoff while meeting QoS requirements in a heterogeneous wireless network, comparing the proposed scheme to TOPSIS. The model in [75] consists of a game between access networks in a converged 4G environment, to decide which service requests should be accommodated by each access network. In [76], the authors study the dynamics of network selection in a heterogeneous wireless environment using evolutionary games. Game theory formulations model decision-making by autonomous independent agents, while in our work we focus on a central global resource controller. Still, the extensive literature on game theory for telecommunications (e.g., [77]) provides rich ideas on network and user utility when considering multiple attributes for optimization.

2.4 Resource Allocation: Implementation Perspective

From an implementation perspective, the interaction between various resource management entities in the hetnet system has to be coordinated through standards-based framework. The resource management entities include GRC, BSs and APs of each RAT in the hetnet system. Also, the level of interaction between these entities and the flexibility in resource allocation decisions depends on the assumptions made about the scheduling mechanism implemented by each RAT in the system. Each cellular RAT employs a flexible scheduler such as deficit weighted round robin, strict-priority, or weighted fair queuing, which can be tuned to achieve various performance objectives such as max-min fairness or proportional fairness or even any other custom objective [78]. On the other hand, Wi-Fi by default has a pre-defined scheduler that implements a Distributed Co-ordination Function (DCF) that employs CSMA/CA with binary exponential backoff algorithm. Moreover, the Wi-Fi AP implements a First-In First-Out (FIFO) queuing system where each arriving packet is served in order. It has been shown that the DCF MAC and FIFO queuing mechanism implemented in a Wi-Fi system leads to equal throughput (max-min fairness) for all associated user devices on a long time-scale [79-80]. Other schemes that obtain proportional fairness objective by achieving airtime fairness in Wi-Fi networks have been proposed [81-83]. But these schemes have not been implemented in practice on a large scale yet. The GRC has to account for these RAT-specific implementation details while performing the user device-to-AP/BS association computations at the global level.

The resource allocation studies for a hetnet system based on a centralized

controller to-date have not modeled these RAT-specific implementation details and the overhead required to set up the centralized control. A few studies try to model the overhead associated with a hetnet system [41-48]. However, these studies deal with the topic of seamless transition between RATs rather than the optimization of resource allocation process. Moreover, such studies employ a trigger-based mechanism where a user device only sends a link parameter report to the centralized controller if its current connectivity condition (usually the RSSI) drops below a certain threshold. In the second phase of our research study, we estimate the overhead required for a centralized solution where each user device in the system sends a link parameter report for all its radio interfaces periodically. We require periodic dissemination of link parameter information to reap the maximum benefit of multi-access network diversity as GRC can make efficient user device-to-AP/BS association decisions more frequently based on each user's time-varying connectivity options.

CHAPTER THREE

SYSTEM DESCRIPTION

3.1 System Model

Figure 3.1 illustrates our generic hetnet system model. The system consists of user devices (also referred to as cognitive user equipment, or cUE) that can connect to one or more AWS. Each AWS comprises of one or more APs or BSs of the same RAT. There are two resource-controlling entities in our proposed system: 1) GRC present in cellular carrier's backend network 2) AP/BS of each RAT. A GRC entity is present in cellular carrier's backend network, providing guidance to **both** the set of independent systems that form the hetnet wireless system **and also** to the reconfigurable devices, instructing them to reconfigure in order for the system to meet global performance objectives or policy requirements. The GRC makes decisions on large time-scales (seconds) using average statistics assigning each user device one or more AP/BS to use for connectivity. The APs/BSs operate over small time-scales (milliseconds) to manage the resources of their corresponding RAT and account for short-term fluctuations in connectivity parameters. The GRC calculates cUE-to-AP/BS mappings (and supported data rate per mapping) and relays the results to each AP/BS as well as the Local Resource Controller (LRC) of each user device. The AP/BS uses this information to establish active connections with the corresponding devices and in making its own scheduling decisions. The LRC uses this information to configure its radios to the specified RAT(s).

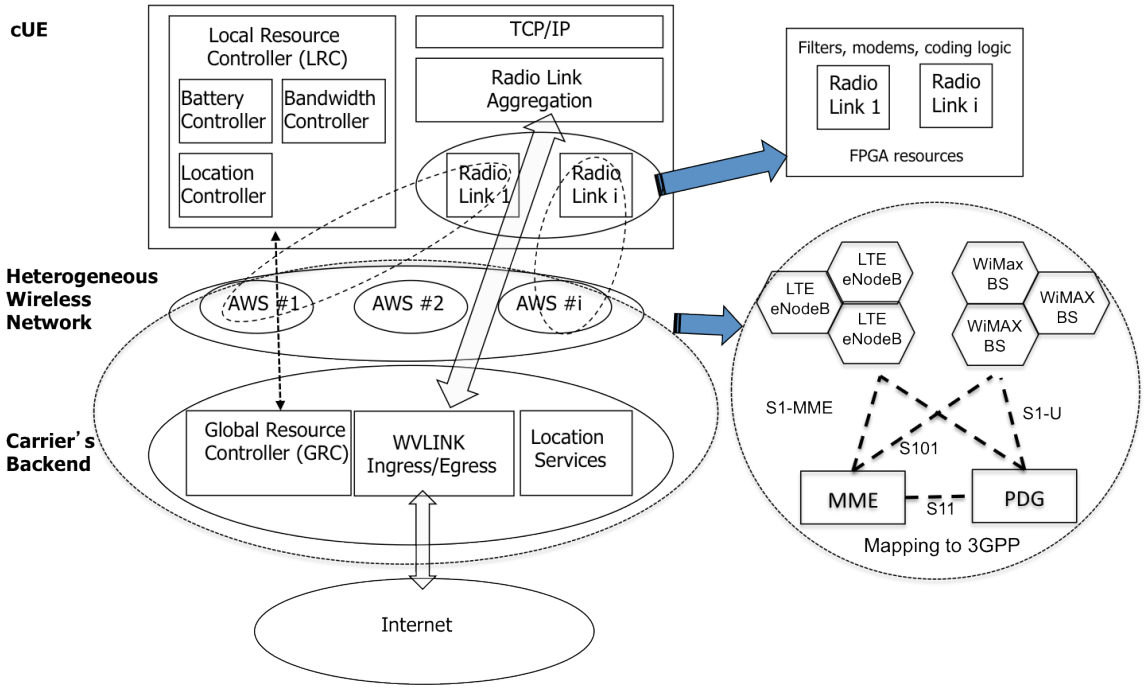

Figure 3.1: System model

The radio link block pictured in Figure 3.1 represents the MAC and physical layers that operate over a portion of the spectrum. A radio is implemented using a combination of custom hardware along with programmable hardware based on technologies such as FPGAs, DSPs, or multi-core ASIPs. User data is tunneled over the unified network cloud via a single ingress/egress point. Entities such as Packet Data Network Gateway (PDG), which includes both Packet Gateway and Signaling Gateway, for 3GPP LTE's System Architecture Evolution (SAE) represent the termination (ingress/egress) point for the tunnel. Additional entities such as Mobility Management Entity (MME) in the carrier's backend network adhere to the functionality described in 3GPP LTE's standards document [84] and help manage information related to each user

in the system while the user is transitioning from one RAT to another. The mapping of our generic system model to the 3GPP LTE's SAE is presented in Figure 3.1.

From an operational perspective, the cUE first must sense for various available RATs and register with the GRC before transmitting any data. We show the procedural flow example of this process in Figure 3.2. First, the cUE senses and scans for available networks and their utilization. Selecting one of the available RATs, the cUE obtains an IP network connection, which it uses to communicate with external hosts. We assume that each user device tries to use the most efficient RAT available initially and follows the following preference order: Wi-Fi, 4G (LTE/WiMAX), 3G (HSPA/EVDO). If the cUE cannot establish a connection to its first preference due to reasons such as very high network load or interference, then it tries to connect to its second preference and this procedure continues until the cUE can establish an initial IP network connection. Next, the cUE discovers, registers, and communicates with the GRC application server, which we assume uses standard discovery and registration procedures as described in [85][1]. After a connection with the GRC is established, the cUE delivers periodic sensing information of available networks to GRC. Upon receiving this periodic sensing information from the cUE, the GRC is able to calculate the cUE-to-AP/BS mappings and the rate assignment per mapping. This mapping information is then relayed to each cUE, which tunes its Reconfigurable Radios (RRs) to the appropriate RATs.

[1] *Registration with an application server involves a combination of DNS lookups with Diameter authentication procedures (RFC 3588), and SIP signaling.*

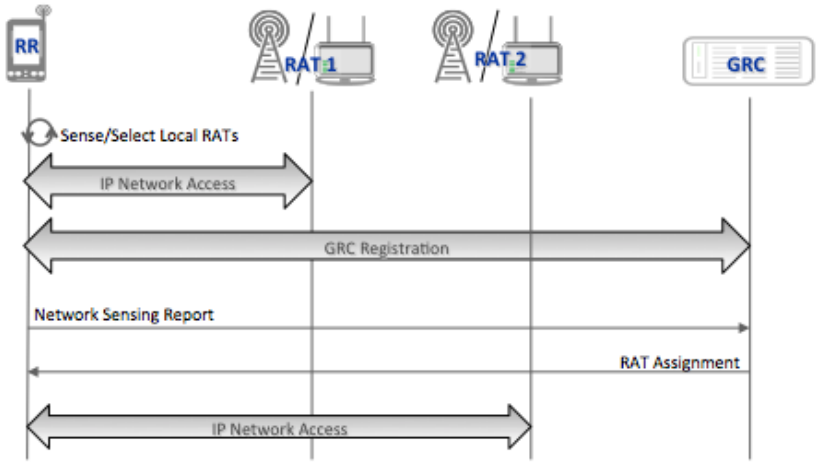

Figure 3.2: Resource allocation procedure

After each RR is configured according to the cUE-to-AP/BS mapping, radio links are established with the associated RATs for data transmission. A pictorial representation of the transmission plane is shown in Figure 3.3. From the perspective of the cUE applications, one TCP/IP stack is managed and scheduled over one or more radio links. The Radio Link Aggregation function is used for packet resequencing and reordering data from each of the RRs. Each RR manages its own radio link and associated protocol with the RATs. User traffic is managed by the GRC (for traffic from the external network to the user device) and by the user device's local resource manager (for traffic flowing from the user device to the external network) on a service flow basis. A service flow has two components: a traffic descriptor and service attributes. The traffic descriptor indicates basic parameters such as allowed sustained flow rates, peak flow rates, etc. The service attributes define the type of service that will be provided including a priority level (other attributes might be defined). Below the Radio Link Aggregation layer, user traffic is managed by the access methods associated with a particular AWS. The Radio Link

Aggregation layer implements functionalities such as traffic splitting/merging for outbound/inbound traffic that is required if multiple radio links are used concurrently. We define the event that corresponds to changing RATs for a particular cUE as a *reconfiguration handoff*. A reconfiguration handoff is a vertical handoff that requires a radio to reconfigure itself. At the hardware level, the cUE will report a cost associated with this reconfiguration handoff. That cost might be either a reduced QoS while the handover is performed or a hard loss of service while switching protocols.

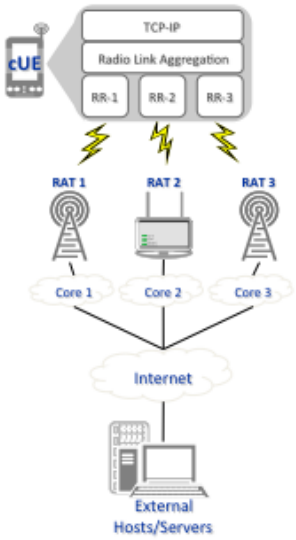

Figure 3.3: Data transmission plane

3.2 Network Co-operation Model

Our approach in the proposed research is based on moving away from the current paradigm of different service providers "locally" optimizing spectrum usage to a new paradigm of "global" spectrum usage optimization. We envision two economic models that could support this move: a *carrier-centric* model and an *Internet* model. In the *carrier-centric* model, a service provider offers services for specific markets where the

cellular carrier might own and operate portions of the physical network and possibly broker 'peering' arrangements with other wireless providers. Customers subscribe to a single cellular operator and gain access to resources or services the subscriber has purchased. An alternative economic model follows the current *Internet model*: organizations own and operate autonomous networks. Unification occurs through an overlay network that can be achieved through a combination of standard protocols, standard services, and incentive/reward mechanisms that promote peering and collaboration. While both are viable models from an engineering standpoint, a successful model has to encompass both economic viability and engineering feasibility. Paradigm shift is likely to be viable only when the technology layer proposed in the research is compatible with the "economic layer" of the network as the actions of network users and providers are driven by economic incentives [86]. However, the economic model is not the focus of our work.

Based on the network co-operation model where cellular providers use peering agreements (which can be based on either carrier-centric or Internet economic model) to allow their subscribers to use other cellular provider's networks, we define two use-cases to increase the coverage and capabilities of a hetnet wireless system. Both use cases assume that two cellular wireless providers (we refer to each as carrier 1 and carrier 2) provide wireless coverage within the same geographic area. The two use cases differ in the level of co-operation that exists between the two carriers. Use case 1 involves x mobile user devices that can connect only to carrier 1's cellular network and x' nomadic user devices that can connect to carrier 1's cellular and Wi-Fi network. Use case 1 also

has y mobile user devices that can connect only to carrier 2's cellular network and y' nomadic user devices that can connect to carrier 2's cellular and Wi-Fi network. Use case 2 allows any mobile user device to make use of the other carrier's cellular network (in addition to it's own carrier's cellular network) and allows any nomadic user device to make use of the other carrier's cellular and Wi-Fi network (in addition to it's own carrier's cellular and Wi-Fi network).

3.3 Performance Measures

The main goal of our proposed centralized solution that operates on a global level is to increase the overall network efficiency due to the benefits of multi-access network diversity. However, network efficiency can be characterized in terms of several conflicting objectives and trying to optimize one objective might result in a very poor performance in terms of another objective. Through our focus on the resource allocation algorithms for a hetnet system in the first phase of our research, we quantify the tradeoffs achievable in terms of four network efficiency performance measures: spectral efficiency, long-term fairness, instantaneous fairness, and overall power consumption. We define each performance measure using the system model terminology given in Table 3.1.

(i) Spectral Efficiency**:** The achievable system spectral efficiency for time interval [t, t+1], denoted γ^t, is represented as the ratio of the (data) rate allocated to each user in the system at time t to the total spectrum used and is presented in (3.1).

$$\gamma^t = \frac{\sum_{u \in U} r_u^t}{\kappa} \tag{3.1}$$

Table 3.1: System Parameters

Parameter	Description
A	Set of BSs/APs for all RATs
U	Set of Users
BU^t	Set of users that are blocked by the admission control procedure at time t
x_{ua}^t	Assignment variable – Determines whether radio $a \in A$ of user $u \in U$ is on or off at time t
r_{ua}^t	Rate (bits/s) allocated to user $u \in U$ by BS/AP $a \in A$ at time t
$r_{ua,max}^t$	Maximum achievable rate (bits/s) for user $u \in U$ through BS/AP $a \in A$ at time t
$r_{ua,norm}^t$	Normalized rate ∈ [0,1] allocated to user $u \in U$ by BS/AP $a \in A$ at time t
r_u^t	Total rate allocated to user $u \in U$ at time t
γ^t	Achievable system spectral efficiency (bits/s/Hz) for time interval [t, t+1]
κ	Total spectrum (Hz) used by the system
η_{ua}^t	Maximum data (in bits) that can be transferred by radio $a \in A$ of user $u \in U$ during time interval [t, t+1]
T^t	Vector containing minimum data rate requirement of each user $u \in U$ to support real-time traffic for time interval [t, t+1]
ω_{ua}^t	Total energy consumed (in Joules) by radio $a \in A$ of user $u \in U$ during time interval [t, t+1]
ω_u^t	Total energy consumed (in Joules) by cUE of user $u \in U$ during time interval [t, t+1]
m_u	Maximum number of usable radios for user $u \in U$ for each time step

The rate allocated to user $u \in U$ at time t, denoted r_u^t, is presented in (3.2) and depends on two parameters: x_{ua}^t - the cUE-to-AP/BS assignment variable at time t, and r_{ua}^t - the rate allocated to user $u \in U$ by AP/BS $a \in A$ at time t.

$$r_u^t = \sum_{a \in A} x_{ua}^t * r_{ua}^t \tag{3.2}$$

Equation (3.2) ensures that the rate allocated to a user only depends on the rate allocated to a user's radio that is currently associated to a RAT by summing the product of x_{ua}^t and r_{ua}^t. Note that r_{ua}^t is a function of the resource blocks assigned to user $u \in U$ by BS/AP $a \in A$ at time t and the supported modulation and coding scheme (MCS). A resource block (RB) is a minimal resource allocation unit. Different RATs use different

terminology when defining a minimal resource allocation unit (for example, Wi-Fi lets users compete for the wireless medium using the CSMA/CA mechanism and lets the contention winner hold the wireless medium for the time necessary to transmit a data frame and ACKs plus any optional control frames associated with virtual carrier sensing; OFMDA-based LTE and WiMAX group twelve consecutive subcarriers in the frequency domain and six or seven symbols in the time domain to form a minimal resource allocation unit). The term for minimal resource allocation unit used by 3GPP based networks (LTE, HSPA) is called a resource block. To unify terminology across all RATs, this term is chosen to represent a minimal resource allocation unit for all RATs in our work.

(ii) Long-Term Fairness: The fairness metric relates to the difference in rates allocated to each user. In general, since best-effort traffic such as FTP has very lenient or no delay constraints, the long-term fairness utility is computed using rates allocated to each user for all time steps under consideration for any study (thousands of seconds). This metric allows resource allocation algorithms to be fair while starving some users for a period of time to take advantage of multiuser diversity where better connected users are given more resources as long as each user eventually gets a fair share of resources (assuming average channel condition for each user in a long run will approximately be the same). We apply a direct mapping of Jain's Fairness Index [87] to compute the long-term fairness metric, ϕ, as presented in (3.3).

$$\phi = \frac{(\sum_{u \epsilon U} \sum_t r_u^t)^2}{|U| * \sum_{u \epsilon U} (\sum_t r_u^t)^2} \tag{3.3}$$

(iii) Instantaneous Fairness: The next generation wireless networks are designed to support user traffic that belongs to several different priority classes. The instantaneous fairness metric applies to high priority traffic classes such as GBR, UGS and rtPS defined by LTE and WiMAX standards, which have minimum data rate requirements per scheduling interval. We assume that the minimum data rate requirement of each user in the system per GRC scheduling interval is represented by the vector $\mathrm{T}^t = [\mathrm{T}_1^t \ldots \mathrm{T}_{|U|}^t]$. For each GRC scheduling interval t, any user $u \in U$ that is allocated enough resources to achieve a data rate of at least T_u^t bits/s can satisfy the needs of his/her real-time applications. To satisfy the real-time traffic demand of as many users as possible, an admission control procedure is required for the resource allocation algorithm. Any user $u \in U$ that cannot achieve a data rate of least T_u^t bits/s for scheduling interval t is considered to be a blocked user and is denoted by BU^t. For each scheduling interval t, the proportion of satisfied users is used to compute the instantaneous fairness utility function, denoted θ_{RT}^t, as described by (3.4). If no users are blocked, the instantaneous fairness metric equals 1 and if all users are blocked, the instantaneous fairness metric equals 0.

$$\theta_{RT}^t = 1 - \frac{|BU^t|}{|U|} \tag{3.4}$$

If support for only best-effort traffic class is assumed where there is no minimum data rate requirement, the instantaneous fairness metric can be computed using Jain's Fairness Index, which is the technique used to compute the long-term fairness metric. However, instead of computing Jain's Fairness Index for rates allocated to each user over large time-scales, Jain's Fairness Index is computed for each scheduling interval t using

the rates allocated to each user for the corresponding scheduling interval to compute the instantaneous fairness metric, θ_{BE}^{t}, as shown in (3.5). The final instantaneous fairness metric for any study is derived by taking an average of the computed Jain's Fairness Index for each scheduling step.

$$\theta_{BE}^{t} = \frac{(\sum_{u \epsilon U} r_u^t)^2}{|U| * \sum_{u \epsilon U} (r_u^t)^2} \tag{3.5}$$

(iv) Power Consumption: The power consumption of a user device depends on two main factors: hardware power consumption and amount of data transfer. For the initial work during the first phase of our research study, we compute the power consumption metric solely based on hardware power consumption costs to focus on differences in overall power consumption for various user device assumptions. For this work, we assume that user devices are equipped with radios that are either static or capable of reconfiguration. Static radios are equipped with one or more non-reconfigurable radios. A non-reconfigurable radio supports a limited level of adaptive capability, but provides the lowest energy consumption due to its custom nature. For example, a non-reconfigurable radio is able to support only one RAT, but it can operate using various MCSs supported by that RAT. A reconfigurable radio is fully adaptive, but consumes comparatively higher amount of power. If a reconfigurable radio moves from the coverage of one RAT to the coverage of another RAT, the GRC will instruct the radio to undergo a reconfiguration handoff, where the radio will reconfigure itself to support the new RAT.

The radios (static or reconfigurable) are made of either ASIC, FPGA or a combination of ASIC and FPGA components. Depending on these components (ASICs or

FPGAs), the power consumption of radios will vary. For static radios, the power consumption is dominated by the dynamic power (P_{dyn}), which is consumed during regular circuit operation. The dynamic power for both ASIC ($P_{dyn,ASIC,a}$) and FPGA ($P_{dyn,FPGA,a}$) based radios for each AP/BS $a \in A$ used in our study is presented in Table 3.2. The dynamic power of FPGAs ($P_{dyn,FPGA,a}$) for each AP/BS $a \in A$ is estimated from [88]. The ratio used for $P_{dyn,FPGA,a}$:$P_{dyn,ASIC,a}$ is 12:1 as recommended by an analysis performed in [89]. For reconfigurable radios, in addition to the dynamic power, there is another source of energy consumption which we label as reconfiguration energy ($E_{rec,FPGA,a}$). $E_{rec,FPGA,a}$ is the energy that is consumed when the circuit of a reconfigurable radio is reconfigured from any AP/BS $b \neq a$ to support AP/BS $a \in A$. The values for $E_{rec,FPGA,a}$ are computed based on the complexity of the RAT standard and the number of blocks that require reconfiguration. We consider only full reconfiguration of radios while computing $P_{rec,FPGA,a}$ values. The minimum reconfigurable block is defined as a data path container (DPC~13.5 KGates). Based on [88], the average reconfiguration power for each DPC (for a Xilinx Virtex II platform) is 234 mW. $E_{rec,ASIC,a}$ represents the increase in energy consumption when ASIC components of radio $a \in A$ are turned 'on' from an 'off' state. We assume that $\mathrm{E_{rec,ASIC,a}}$ is almost negligible compared to $E_{rec,FPGA,a}$ and that $E_{rec,ASIC,a}$ for Wi-Fi chipsets is much lower than that for cellular chipsets. The actual $E_{rec,FPGA,a}$ and $E_{rec,ASIC,a}$ numbers used in this study are presented in Table 3.2.

Table 3.2: Hardware implementation and power consumption statistics for current technologies

	802.11g	802.16e	LTE	HSDPA	EVDO
No. of Kgates	416	728	270	723	684
No. of DPCs	31	53	20	53	50
$P_{dyn,FPGA,a}$ (Watts)	1.76	3	1.13	3	2.83
$P_{dyn,ASIC,a}$ (Watts)	0.15	0.25	0.09	0.25	0.24
$E_{rec,FPGA,a}$ (Joules)	7.25	12.4	4.68	12.4	11.7
$E_{rec,ASIC,a}$ (Joules)	0.05	0.28	0.28	0.64	0.64

To determine the portion of the radio implemented using FPGA and ASIC technology, we define a scalar $\beta \in [0,1]$ that represents the percentage of radio components manufactured using FPGA technology. The percentage of radio components manufactured using ASIC technology is $1-\beta$. In addition, we use another scalar, $\lambda \in [0,1]$, which we define as *impact of reconfiguration*, to capture the effects of technical improvements in radio systems. $\lambda = 1$ represents using reconfiguration energy costs presented in Table 3.2. But as hardware evolves (or concepts such as partial reconfiguration gain momentum), these costs will go down. As a result of innovative radio design architectures, the reconfiguration energy consumption will only be a fraction of that presented in Table 3.2. The scalar λ captures this effect and as $\lambda \to 0$, the reconfiguration energy cost becomes almost negligible. Using these scalars and the power/energy consumption values presented in Table 3.2, the total power consumption

metric for our hetnet system can be calculated using (3.6), where $\alpha_{run,ua}$ and $\varphi_{rec,ua}$ represent the percentage of (simulation) time user $u \in U$ spends being connected to AP/BS $a \in A$ and the average number of reconfigurations (per second) user $u \in U$ experiences to support AP/BS $a \in A$ respectively.

$$P_{total} = \sum_{u \in U} \sum_{a \in A} \{\alpha_{run,ua}[\beta . P_{dyn,FPGA,a} + (1 - \beta) . P_{dyn,ASIC,a}] + \varphi_{rec,ua}[\lambda . E_{rec,FPGA,a} + (1 - \lambda) . E_{rec,ASIC,a}]\} \quad (3.6)$$

For the latter parts of our research study where we limit user device assumptions, we update our metric to incorporate both the energy consumption due to data transfer and due to hardware reconfigurations. From a RAT protocol standpoint, various schemes to transition into energy-efficient modes (such as deep sleep mode) have been developed when the radio is not transmitting/receiving any data traffic [90-92]. In deep sleep mode, the radios turn off most of the circuitry and hence consume negligible amounts of energy. As a result, we move to a model that is based on the amount of data transferred (transmitted/received) by a radio and remove the dynamic power component from the model, which assumes the radio circuitry always remains in a 'normal' power consuming state while it is connected to any RAT. We use a linear energy consumption model, which is similar to the model proposed in [93-94]. The energy consumption for user $u \in U$ during time interval [t, t+1], denoted as ω_u^t, is computed using (3.7). The first energy consumption component, $E_{t,a}$, relates to the transfer energy component described in [94] and it depends on η_{ua}^t, the maximum number of data bytes that can be transferred by radio $a \in A$ of user $u \in U$ during time interval [t, t+1]. The second energy component,

$E_{o,a}$, represents the overhead energy incurred during a reconfiguration handoff and has two sub-components. The first sub-component, $E_{rec,a}$, represents the extra energy that is spent by RRs in reconfiguring the hardware to transition to a new RAT. We assume an FPGA platform as our RR platform [95] and use energy consumption numbers represented by $E_{rec,FPGA,a}$ in Table 3.2. The second sub-component, $E_{assoc,a}$, represents the extra energy that is spent associating with a new RAT and is similar to the ramp energy concept used in [94]. We summarize the energy consumption numbers for all the components in the new energy consumption model in Table 3.3. The overall energy consumption metric for each user for the entire simulation duration, ω_u, is computed by summing the energy consumption of the user computed for each time interval [t, t+1] using (3.7). The computed ω_u value for each user is summed and the sum is divided by the simulation duration to obtain average power consumption per user.

$$\omega_u^t = \sum_{a \epsilon A} [\, E_{t,a}(\eta_{ua}^t) + (x_{ua}^t - x_{ua}^{t-1}) * (1 - x_{ua}^{t-1}) * E_{o,a}] \tag{3.7}$$

Table 3.3: Energy consumption components for current technologies

	802.11g	802.16e	LTE	HSDPA	EVDO
$E_{t,a}$ (Joules/x KB)	0.007(x)	0.018(x)	0.018(x)	0.025(x)	0.025(x)
$E_{rec,a}$ (Joules)	7.25	12.4	4.68	12.4	11.7
$E_{assoc,a}$ (Joules)	5.9	3.2	3.2	3.5	3.5
$E_{o,a}$ (Joules)	13.15	15.6	7.88	15.9	15.2

CHAPTER FOUR

RESOURCE ALLOCATION

For the first phase of our research study, we explore the resource allocation procedure implemented by the centralized GRC and study the tradeoffs surrounding network efficiency measures of spectral efficiency, long-term fairness, instantaneous fairness, and overall power consumption. Within the first phase of study, we first analyze the achievable tradeoffs in terms of network efficiency measures of spectral efficiency and power consumption based on different user device assumptions, network topologies and network outages. We then perform an optimization study in terms of all four network efficiency measures where we use a utility function-based and a weighted sum approach to optimize system performance in terms of all four network performance measures using a two-step resource allocation procedure. These two studies are the focus of this chapter and are presented in detail next.

4.1 Problem Assumptions

The hetnet system that we consider for this phase of work consists of Wi-Fi, LTE, WiMAX, HSPA and EVDO RATs, GRC, and reconfigurable (or multi-modal) user devices. We use a high-level system modeling approach to perform this phase of our research study. In doing so, we make the following assumptions:

(i) Flexible Scheduler implementation at each RAT: The GRC computes user device-to-BS/AP association decision and the supported data rate per association mapping each scheduling interval. The association information is used by the user device to tune its radio to the corresponding RAT, and the data rate per association mapping is

used by the scheduler implemented at the AP/BS of each corresponding RAT to allocate appropriate amount of RBs to each connected user device. For this solution to be feasible, we use the underlying assumption that each RAT implements a flexible scheduler that can control the amount of resources allocated to each connected user device. For cellular RATs, this assumption is easy to incorporate in a real system as cellular systems are controlled by a centralized BS that implements a flexible scheduler such as deficit weighted round robin, strict-priority, or weighted fair queuing. Setting the weights of each queue to appropriate values ensures appropriate distribution of RBs to each connected user devices. For Wi-Fi RATs, this assumption is challenging to implement in a real system as Wi-Fi uses the distributed CSMA/CA scheduling. Extensions to the base 802.11g protocol, such as 802.11e, have made it possible to provide four different levels of priority (and throughput) to user devices, but still fine-grained control required by our approach is not yet available in current Wi-Fi solutions. However, several studies have proposed the use of separate queues for each connected user device at the AP. Setting the congestion window (CW_{min} and CW_{max}) parameters for each queue appropriately results in the fine-grain control for RB distribution to each user device required by our solution [81,83]. We assume this functionality exists in the Wi-Fi APs used in our proposed solution.

(ii) 25% overhead for each RAT for supporting messaging framework required for a centralized solution: Each RAT in our system uses an adaptive MCS. The signal strength achieved by various radios on a user device (which is based on the distance of the user device from the corresponding BS/AP) dictates the MCS used by the radios on

the device to connect to the corresponding BS/AP. The MCS dictates the maximum achievable data rate ($r^{t}_{ua,max}$ parameter presented in Table 3.1) for each radio on each user device by determining the maximum number of bits that can be transmitted over each RB. However, to account for overhead required for supporting messaging framework in our proposed centralized solution, we deduct 25% RBs from each RAT, which reduces the maximum achievable data rate from suggested theoretical maximum data rates by 25% for each RAT. The details on maximum achievable data rates for each RAT in our study are presented in Tables A1-A5 in Appendix A.

(iii) Each user device has three reconfigurable radios that can be used concurrently: The number of radios equipped on a user device keep increasing with time as space and energy-efficient hardware architectures are constantly innovated due to Moore's law. Usually, a user device today is equipped with at least a Wi-Fi radio and a cellular radio. The cellular technologies that are deployed in practice are based either on GSM (HSPA) or CDMA (EVDO) standard with the upcoming technologies such as LTE/LTE-Advanced and WiMAX/WiMAX-Advanced moving to a flat all-IP architecture. To be able to connect to each RAT that has been deployed in practice, we assume each user device is equipped with three reconfigurable (or multi-modal) radios. Moreover, in our study we consider the *fractional association* scenario where each user device can simultaneously use multiple radios to support various application data flows, and traffic from each data flow can be split over these radios in an intelligent manner. The fractional association scenario represents a more futuristic vision and clearly provides a better solution in terms of optimality than the *integral association* scenario

used in practice today, where only one radio can be used at any given time. The 3GPP frameworks such as Joint Radio Resource Management (JRRM), which defines traffic splitting service, and recent work related to multihoming capability using IETF protocols such as SCTP [96] indicate a strong interest for future support of fractional association scenario. We assume the use of such capabilities in our solution to support the fractional association scenario.

(iv) The GRC operates on a one-second scheduling interval: The intent of our proposed solution is to let GRC make periodic decisions on large time scales (seconds or minutes), while the BSs/APs of each RAT make scheduling decisions on small time scales (milliseconds) to account for short-term fluctuations in connectivity conditions. While some of the settings are customizable for LTE and WiMAX, generally these RATs generate a schedule every 5 or 10 milliseconds. HSPA typically generates a schedule every 2 milliseconds and EVDO generates a schedule every 26.67 milliseconds. Wi-Fi typically assigns a channel to the user for 0.5 milliseconds to send one data frame (which includes the DIFS, Data, SIFS, ACK mechanism). The GRC performs global-level optimization (re-associations) and has to operate on larger-time scales to account for issues such as overhead/result propagation delay. So, to minimize actual overhead and to make sure that the user devices and BSs/APs of various RATs can use the decisions made by the GRC, a scheduling interval of 1 second is used for the GRC in our study.

(v) Infinitely backlogged downlink data traffic: We consider data traffic flow in the downlink direction (from BS/AP to user device). For both best-effort and real-time traffic, we assume that the data connection queues supporting each traffic type for each

user device at the BS/AP are always fully backlogged. So all the resources allocated to each user device by the BS/AP are fully utilized.

Using these five assumptions, we conduct two simulation studies. From a resource allocation standpoint, the first study is based on heuristic algorithm that considers all four network efficiency measures of spectral efficiency, instantaneous and long-term fairness and overall power consumption. The algorithm tries to achieve a balance in performance related to all these efficiency measures. However, benefits of network co-operation and tradeoffs achieved in terms of system spectral efficiency and overall power consumption based on different user device assumptions are the main objectives of this study. The second study is based on an optimization algorithm that achieves a balance in tradeoffs in terms of all four network efficiency performance measures. For this work, we assume that the reconfigurable radios present at user devices are fabricated using FPGA platform and we analyze the performance results in terms of all four network efficiency performance measures.

4.2 Heuristic Algorithm Simulation Study

We consider the presence of two major cellular carriers in a 2 * 2 km^2 area that operate multiple RATs. We use EVDO (3G), HSPA (3G), WiMAX (4G), LTE (4G) and IEEE 802.11g (Wi-Fi) in our experiments as the representative RATs that current cellular carriers support. The 3G base stations (EVDO, HSPA) have a coverage radius of 1.50 km. The 4G base stations (WiMAX, LTE) have a coverage radius of 1.0 km. The Wi-Fi APs have a coverage radius of 0.15 km. Network planning (the AP/BS location of each RAT in our network topology) has a significant impact on the achieved spectral

efficiency. If each carrier has a similar amount of network resources in a given area, the data rate allocated to the users of each cellular carrier is almost equal. But if one carrier has more network resources than the other carrier, then the data rate allocated to the users of the first carrier is much greater than the users of the second carrier. The impacts of sharing resources across carriers for these equal and unequal carrier resource scenarios will vary significantly. To study the effects of such differences in each operator's network resources, we create two network deployment scenarios: 1) *Balanced Network Topology* 2) *Unbalanced Network Topology* shown in Figures 4.1 and 4.2 respectively.

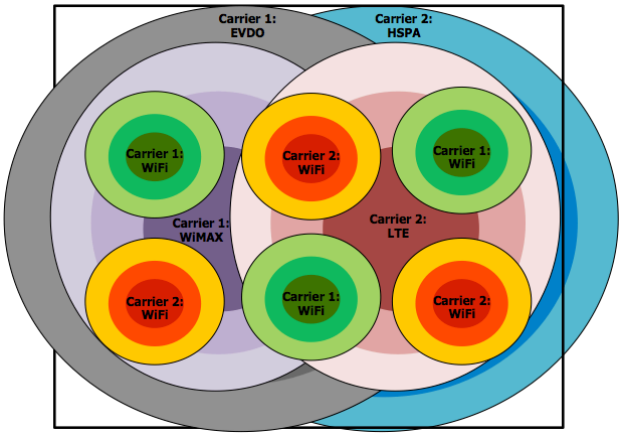

Figure 4.1: Balanced Network Topology

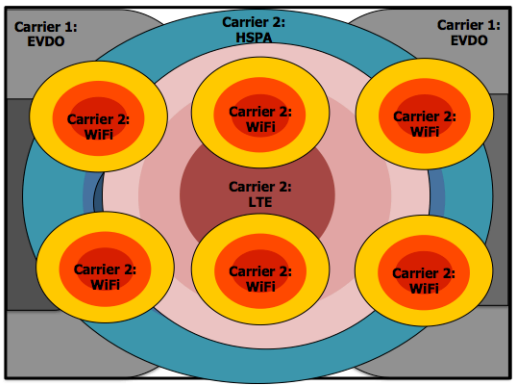

Figure 4.2: Unbalanced Network Topology

In our balanced network topology, each cellular carrier deploys a 3G technology (EVDO – carrier 1, HSPA – carrier 2), a 4G technology (WiMAX – carrier 1, LTE – carrier 2) and 3 Wi-Fi APs in the 2 * 2 km^2 grid. The 3G/4G base-stations are placed very close to the center of the grid and the Wi-Fi APs are spread throughout the topology to give each carrier equal network coverage. For our unbalanced network topology, carrier 1 only has 3G network coverage (EVDO) from two base-stations placed at the two horizontal edges of the grid, whereas carrier 2 has network coverage from one 3G base-station (HSPA) and one 4G base-station (LTE) placed at the center of the grid and six Wi-Fi APs that are spread throughout the topology. In the network topologies shown in Figures 4.1 and 4.2, the different color shades represent various MCS levels supported by each RAT. The darker the shade, the higher the MCS a radio can use in a specific area. The actual MCSs for each RAT are presented in Tables A1-A5 in the Appendix.

In addition to the two network topologies, we study two use cases described in the network co-operation model in Chapter 3. Use case 1 involves mobile user devices that can connect only to its own carrier's cellular network and nomadic user devices that can connect only to its own carrier's cellular and Wi-Fi network. Use case 2 allows any mobile user device to make use of the other carrier's cellular network and any nomadic user device to make use of the other carrier's cellular and Wi-Fi network. Furthermore, for use case 1, since a user device can only connect using RATs of its own carrier, the device is equipped with three static radios which supports each device's corresponding carrier's RATs. So the user devices assumed for use case 1 are static multi-modal devices. For use case 2, each user device is equipped with three reconfigurable radios and is capable of

supporting all five RATs presented in the network topology. So the user devices assumed for use case 2 are reconfigurable devices.

Both balanced and unbalanced network topology presented in Figures 4.1 and 4.2 respectively is used as the simulation topology. The simulation involves 100 user devices (or nodes), 50 of which are subscribed to carrier 1 and the other 50 subscribed to carrier 2. For the balanced topology, 75 percent of the total nodes are mobile nodes and the remaining 25 percent are nomadic nodes, and the mobile and nomadic nodes are split evenly between both carriers. So, 37 nodes for carrier 1 are mobile nodes and 13 of them are nomadic nodes. 38 nodes for carrier 2 are mobile nodes and 12 of them are nomadic nodes. Mobile nodes are allowed to move freely in the entire 2 * 2 km^2 grid, whereas the nomadic nodes are confined to move in an inner 1 * 1 km^2 grid that encompasses all Wi-Fi APs. Mobile nodes move using a random waypoint mobility model at a constant speed of 20 mph. Nomadic nodes move using a random waypoint mobility model at a constant speed of 2 mph. For the unbalanced topology, all 100 nodes (50 subscribed to carrier 1 and the other 50 subscribed to carrier 2) are nomadic nodes. The nodes of carrier 1 are clustered in 2 groups. The first group is located on the left side of the grid (centered at [0 m, 1000 m]) and the second group is located on the right side of the grid (centered at [2000 m, 1000 m]). The nodes of carrier 2 form the third cluster and are located in the center of the grid (centered at [1000 m, 1000 m]). All nodes are allowed to move at speeds of 2 mph in a restricted space of 500*500 m^2 using random waypoint mobility model based on the cluster they belong to. This leaves each user with relatively bad coverage for

use case 1 and significantly improves performance when they start using resources of the other carrier under policies of use case 2.

Each node uses radios according to the decisions made by the GRC. When the GRC instructs a node to switch/reconfigure the radio to be used, there is a cost associated with this operation in terms of temporary downtime and an increase in energy consumption. Because the GRC scheduler operates on a 1 second allocation basis, we approximate the communication downtime cost by not allocating any bandwidth to the radio for 1 second. If we assume that communication downtime includes hardware reconfiguration times (for reconfigurable devices) and the time required to establish the new physical and logical link connections with a new RAT, a downtime cost of 1 second seems reasonable based on vertical handover times claimed to be between few hundred milliseconds to a few seconds by the work presented in [33]. The energy consumption cost during reconfiguration is presented in Table 3.2. Since both communication downtime and increase in energy consumption are hardware and implementation specific, we multiply the communication downtime (1 second) and reconfiguration energy cost (Table 3.2) with the impact of reconfiguration ($\lambda \in [0,1]$) experimental parameter.

We do not include a detailed channel model in our studies, but rather introduce an artificial degradation in network quality. We use a parameter which we refer to as *network outage* to model the percentage of time any network is unavailable to the users. An outage might occur as a result of a number of situations including congestion due to increased network load, increased RF interference levels, AP/BS malfunction/software upgrades, or even network attacks such as denial of service. The network outage is an experimental

parameter that controls the percentage of RBs of a AP/BS that are effectively not used. The outage percentage ranges from 0% to 25% in increments of 5% in our simulation. Each AP/BS suffers independent random outages with the probability determined by the network outage percentage.

4.2.1 Heuristic Resource Allocation Algorithm

The GRC implements a sort-based scheduler that assigns each user device the most efficient access technology and that allocates bandwidth in a manner which seeks fairness while maximizing achievable system throughput. Support for both real-time and best-effort traffic expected to be an integral part of future wireless hetnet systems is assumed. To satisfy the real-time traffic requirements in addition to providing best-effort service, we develop a two-step heuristic algorithm that attempts to satisfy the minimum data rate requirements (T_u^t bits/s) of each user per scheduling interval in the first step and allocates the remaining resources to the users that can make the best use of those resources in the second step. We assume the same instantaneous data rate requirement (100 kbps) for each user in the system, i.e. T_u^t = 100 kbps for $\forall u \in U$. The pseudo-code for our algorithm is presented in Appendix B.

Since the scheduler implemented at Wi-Fi APs is not very flexible, the GRC algorithm assigns resources for Wi-Fi and the cellular RATs in a separate manner. For assigning Wi-Fi resources, the algorithm checks the number of nomadic users that are in range of a Wi-Fi AP. It assigns equal number of Wi-Fi RBs to all nomadic users that can connect using a particular Wi-Fi AP by dividing the total number of RBs the AP possesses by the total number of users that can connect to it. This procedure generates a schedule

that achieves proportional fairness at each Wi-Fi AP. For assigning cellular RAT resources, the algorithm follows a two-step approach. In the first step, the algorithm allocates a data rate of 100 kbps (represented as T_u^t in the pseudo-code) to each node using its best cellular (3G/4G) radios (based on the sorted order of radios for each node in terms of MCS). In the second step, the scheduler distributes unused cellular access technology resources to a window (ζ) of 10 mobile/nomadic nodes with best connectivity parameters (based on sorted order of radios for each RAT in terms of MCS) in increments of 100 kbps (represented as ε in the pseudo-code). The overall order of allocation follows technologies that can achieve the highest theoretical data rate to the technologies that can achieve the lowest theoretical data rate. So, the scheduler assigns resources in the following order: Wi-Fi, 4G (LTE, WiMAX) and then 3G (HSPA, EVDO) technologies. All the nodes are limited to a maximum allocation of 1 Mbps during the cellular technology allocation phase. Any node that reaches 1 Mbps or is already above 1 Mbps (for example, any nomadic node that was assigned more than 1 Mbps by Wi-Fi) is not assigned any additional resources. The scheduler implementation is intended to be a simple heuristic algorithm that accounts for instantaneous fairness (for real-time traffic) and provides performance close to a proportional fairness objective with more bias towards spectral efficiency compared to long-term fairness for best-effort traffic.

4.2.2 Heuristic Algorithm Results and Analysis

Each simulation is run for 10,000 seconds in MATLAB. The results from the simulations include average spectral efficiency and the average power consumption per node that are observed as the two experimental parameters (network outage and the

relative impact of reconfiguration) are varied. We compute the spectral efficiency (in bits/sec/Hz) for each scheduling interval according to (3.1). At the end of a simulation run, we average the spectral efficiency computed for each scheduling interval to derive the average spectral efficiency. The total power consumption of each node (in Watts) is calculated using (3.6). At the end of the simulation, the aggregate power consumption of all nodes is divided by the number of nodes resulting in the average power consumption per node.

4.2.2.1 Spectral Efficiency Results

4.2.2.1.1 Balanced Network Topology

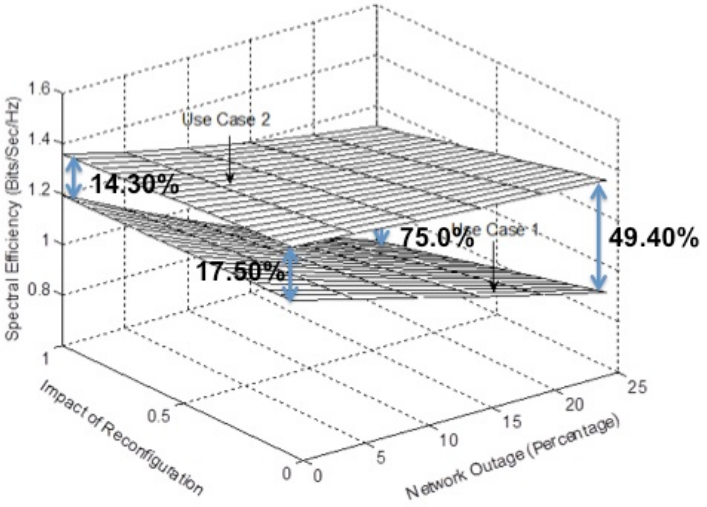

Figure 4.3: Spectral Efficiency for Balanced Topology

The spectral efficiency for the balanced network topology is presented in Figure 4.3. As expected, use case 2 utilizes the spectrum more efficiently than use case 1. Reconfiguration allows the global and local controllers to assign each node to the most

efficient APs/BSs. To provide lower bounds, for no network outage and impact of reconfiguration of 1, the spectral efficiency gain for use case 2 (1.36 bits/sec/Hz) when compared to use case 1 (1.19 bits/sec/Hz) is approximately 14.30%. The spectral efficiency decreases as the network outage increases as can be observed from Figure 4.3. This phenomenon is intuitive since network outage results in loss of resources that could have been used to allocate a higher data rate to each user. Also as expected, the rate of decline for use case 1 where there is no carrier collaboration (static radios) is much steeper than use case 2 where carrier collaboration (reconfigurable radios) does exist as the experimental parameters, network outage and impact of reconfiguration, increase. The maximum spectral efficiency gain for use case 2 (1.12 bits/sec/Hz) when compared to use case 1 (0.64 bits/sec/Hz) is around 75.0% when there is 25% network outage and the impact of reconfiguration is 1. This highest gain of 75.0% is limited by the balanced network topology where both carriers have almost equal amount of resources.

4.2.2.1.2 Unbalanced Network Topology

The spectral efficiency for unbalanced network topology is presented in Figure 4.4. Again as expected, reconfiguration allows use case 2 to utilize the spectrum more efficiently than use case 1. To get a lower bound, when there is 25% network outage and impact of reconfiguration of 0, the spectral efficiency gain for use case 2 (1.43 bits/sec/Hz) when compared to use case 1 (0.34 bits/sec/Hz) is around 314.3%. The maximum spectral efficiency gain for use case 2 (1.79 bits/sec/Hz) when compared to use case 1 (0.27 bits/sec/Hz) is around 553.7% when there is no network outage and the impact of reconfiguration is 1. The increase in spectral efficiency range [314.3%,

553.7%] is quite high for unbalanced network topology when compared to a balanced network topology [14.3%, 75.0%]. This phenomenon results due to the fact that in the unbalanced topology, for use case 1 all the users connect to APs/BSs supported by their own carrier at very low data rates since they are at the edge of their carrier's network coverage. But for use case 2 when all the users can connect to any available AP/BS, they connect with APs/BSs supported by the other carrier at very high data rates since they are very close to those APs/BSs. This shows the tremendous gains that are possible in a realistic unbalanced network deployment scenario where resources of one carrier exceed those of another if a truly heterogeneous wireless system is created where all available resources in a given area are managed at a global level.

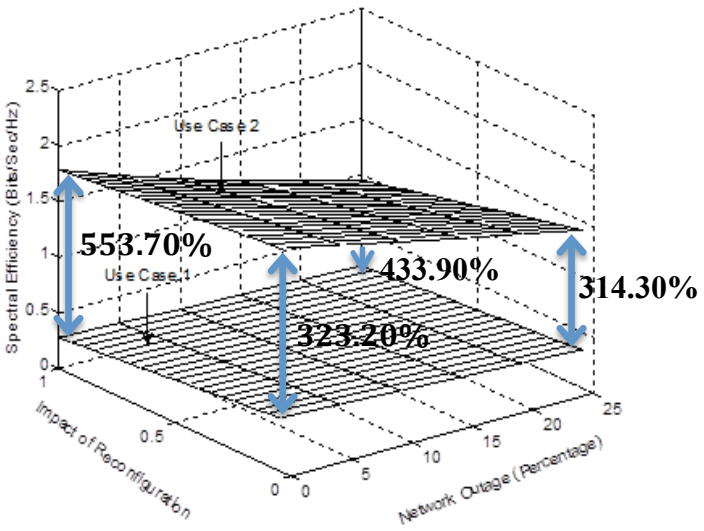

Figure 4.4: Spectral efficiency for unbalanced topology

4.2.2.2 Power Consumption Results

Power consumption in our heterogeneous wireless system for this study depends on two factors: the number of reconfigurations and the type of hardware fabric used (ASIC/FPGA). The number of reconfigurations depend on the number of connectivity options for each node when GRC comes up with node-AP/BS mapping every scheduling period. Since nodes have more connectivity options under use case 2, there are a greater number of reconfigurations for use case 2 compared to use case 1. A reconfiguration for use case 1 is equivalent to switching one radio *off* and turning another radio *on*. For use case 2, a reconfiguration requires the circuitry of one radio to be switched to support a different RAT. From the hardware perspective, since we assume static multi-modal radios for use case 1, the radios are made up of complete low-power consuming ASIC components, or using our power consumption model described in (3.6), $\beta = 0$. For use case 2, since the radios require reconfigurable components, we investigate three hardware settings: i) radio is made up of completely FPGA components, i.e. $\beta = 1$ ii) radio is made up of 50% FPGA and 50% ASIC components, i.e, $\beta = 0.5$ iii) radio is made up of completely ASIC components, i.e. $\beta = 0$.

4.2.2.2.1 Balanced Network Topology

The result for the first hardware setting for balanced network topology is provided in Figure 4.5. As shown, the increase in power consumption lies in the range [114.0%, 916.8%] when the radios are implemented completely using FPGA fabric as compared to a complete ASIC implementation. The highest increase in power consumption occurs when the impact of reconfiguration is 0. This suggests that for a balanced network

topology, the hardware choice has a greater impact than the number of reconfigurations on average power consumption. While the power consumption does increase as the impact of reconfiguration increases, the relative difference between the two use cases becomes smaller. The same phenomenon is observed for second hardware setting when 50% of the radio fabric is made using ASIC components and the other 50% is made up of FPGA components as seen in Figure 4.6. However, the low-energy consuming ASIC components decrease the power consumption by almost half and now the increase in power consumption lies in the range [73.1%, 486.7%]. The third hardware setting is not feasible today in building a completely reconfigurable device, but is studied to provide an intuition on how much extra power is consumed if the only difference between the two use cases is the number of connectivity options available to each node per scheduling interval. As can be seen from Figure 4.7, the increase in power consumption lies in the range [32.2%, 129.8%].

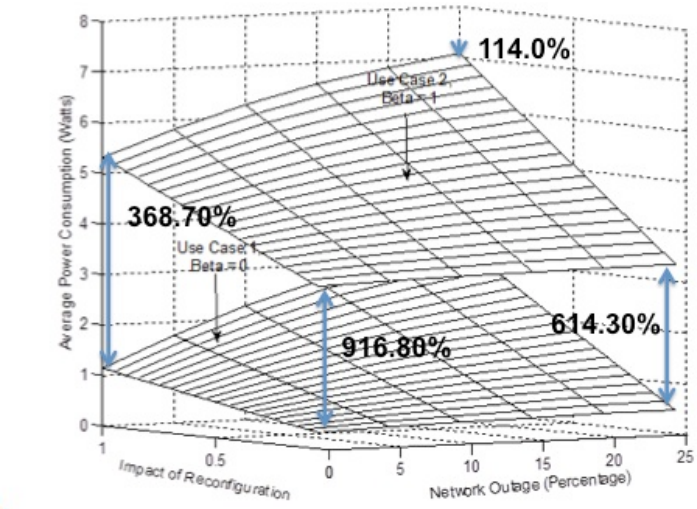

Figure 4.5: Average power consumption for ASIC (Use Case 1, Beta = 0) vs. FPGA (Use Case 2, Beta = 1) implementation for balanced network topology

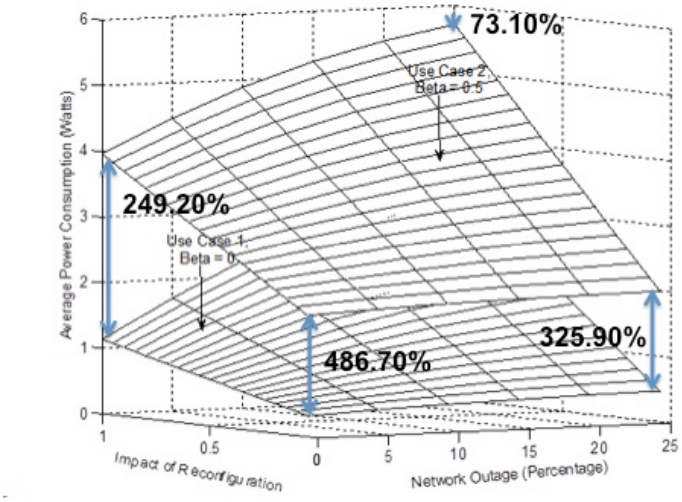

Figure 4.6: Average power consumption for ASIC (Use Case 1, Beta = 0) vs. 50% ASIC, 50% FPGA (Use Case 2, Beta = 0.5) implementation for balanced network topology

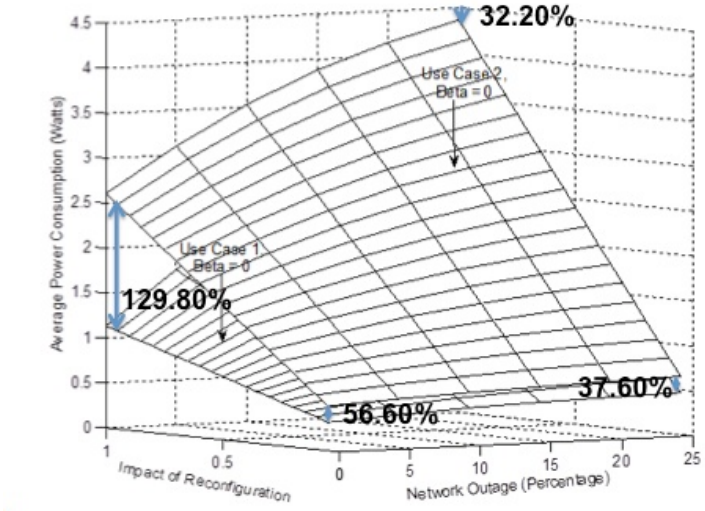

Figure 4.7: Average power consumption for ASIC (Use Case 1, Beta = 0) vs. ASIC (Use Case 2, Beta = 0) implementation for balanced network topology

The reconfiguration rate for the balanced network topology is presented as a function of network outage probability in Figure 4.8. Since each node has three radios in the simulation, the reconfiguration rate values can range between [0, 3]. From Figure 4.8, we see that the actual values of reconfiguration rate lie between 0.20 and 1.5. For smaller network outage percentage, the reconfiguration rate for use case 2 is much higher in

comparison to use case 1. This is justified since more reconfigurations are performed because better resources become available to nodes as they move according to their movement pattern and not because of the network outage. Network outage has lesser effect than the number of available resources in this case. Since nodes in use case 2 have access to more resources, these nodes experience a greater level of reconfiguration than use case 1 nodes. But as the network outage approaches 25%, the difference between reconfigurations for use case 1 and use case 2 decreases. This result helps explain the power consumption trend seen in Figures 4.5-4.7. As the network outage increases, the difference in number of reconfiguration between two use cases decreases and as a result the difference in power consumption decreases.

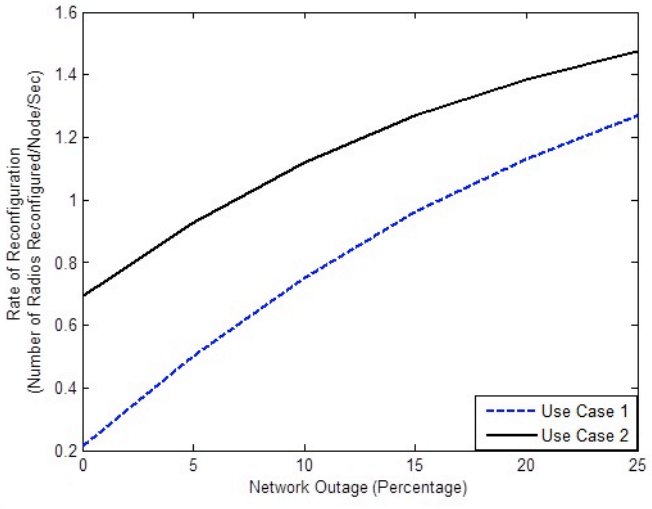

Figure 4.8: Reconfiguration rate for balanced network topology

4.2.2.2.2 Unbalanced Network Topology

The result for first hardware setting for unbalanced topology is provided in Figure 4.9. As can be seen from the figure, the increase in power consumption lies in the range

[104.9%, 614.9%] when the radios are implemented completely using FPGA fabric as compared to a complete ASIC implementation. The increase in power consumption of over 600% might be too costly even though the gain in spectral efficiency is about 550% for that setting. As an alternative, when 50% of the radio fabric is made using ASIC components, the increase in power consumption lies in the [70.0%, 355.4%] range as shown in Figure 4.10. The increase in spectral efficiency of about 550% at the cost of increase in power consumption of about 350% would be a better choice to implement reconfigurable radios. The average number of radios used per node for use case 2 is 1.23. At each time step, none of the nodes use more than 2 radios. So, it might suffice to limit the number of reconfigurable radios per node implemented using FPGA fabric and have some static radios that use low-power custom built circuitry (ASIC fabric). How much of this hybrid architecture is possible today is still an open question and is currently being investigated by several researchers. The results of the infeasible third hardware setting are studied to provide an intuition on how much extra power is consumed if the only difference between the two use cases is the number of connectivity options available to each node per scheduling interval. As can be seen from Figure 4.11, the increase in power consumption lies in the range [35.1%, 98.8%]. So just based on an increase in number of connectivity options, which results in a higher rate of reconfiguration for use case 2, and using the same hardware components in constructing radios for both use cases results in a huge increase in spectral efficiency (553.7% for unbalanced topology) at twice the amount of power consumption.

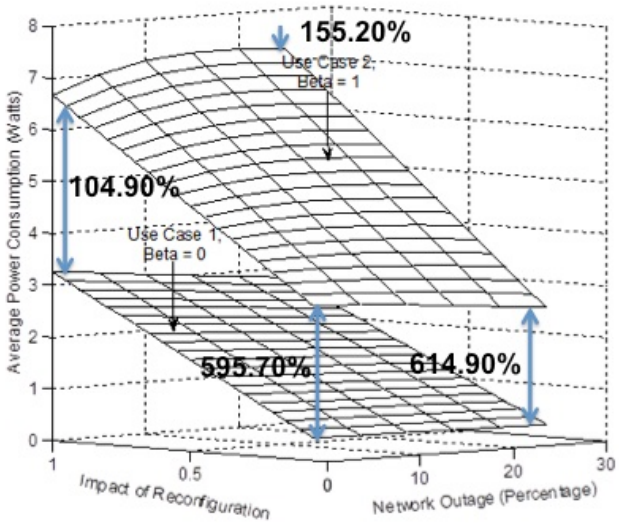

Figure 4.9: Average power consumption for ASIC (Use Case 1, Beta = 0) vs. FPGA (Use Case 2, Beta = 1) implementation for unbalanced network topology

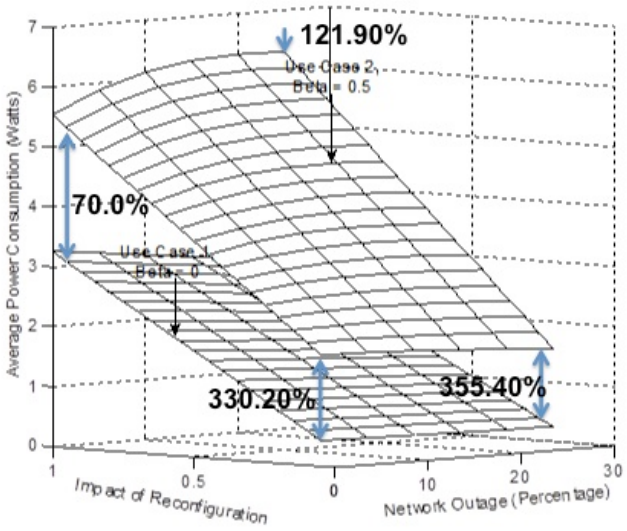

Figure 4.10: Average power consumption for ASIC (Use Case 1, Beta = 0) vs. 50 % ASIC, 50% FPGA (Use Case 2, Beta = 0.5) implementation for unbalanced network topology

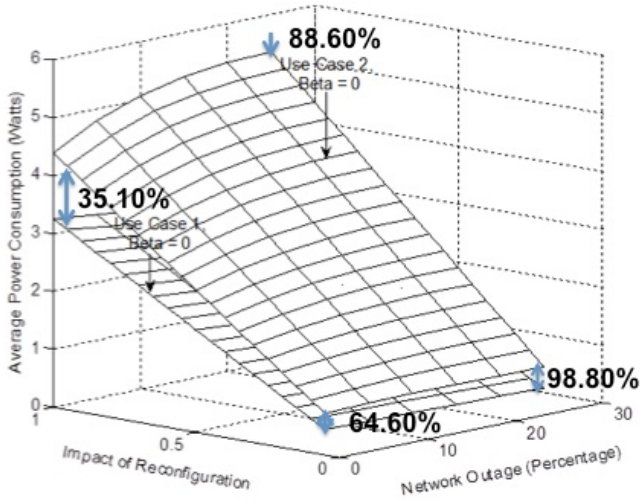

Figure 4.11: Average power consumption for ASIC (Use Case 1, Beta = 0) vs. ASIC (Use Case 2, Beta = 0) implementation for unbalanced network topology

To provide an understanding of the trends seen in Figures 4.9-4.11, the rate of reconfiguration for the unbalanced network topology is presented in Figure 4.12. Due to the restricted movement pattern in the unbalanced network topology scenario, users in use case 1 have a hard time getting resources compared to use case 2. So use case 1 actually has more reconfigurations than use case 2 when there is no network outage. But as the network outage increases and reaches 20%, the number of reconfigurations for use case 2 approaches those of use case 1 and eventually surpasses them. As a result, the difference in power consumption between the two use cases increases as the network outage increases. In addition, resources of two of the six Wi-Fi APs are not employed for use case 1 whereas they are utilized for use case 2. So the actual power consumption not only depends on the rate of reconfiguration, but also the number of APs/BSs that are used. Due to the usage of two extra APs, the power consumption of use case 2 is always

greater than that of use case 1 (even when same ASIC hardware is used for radios for both use cases).

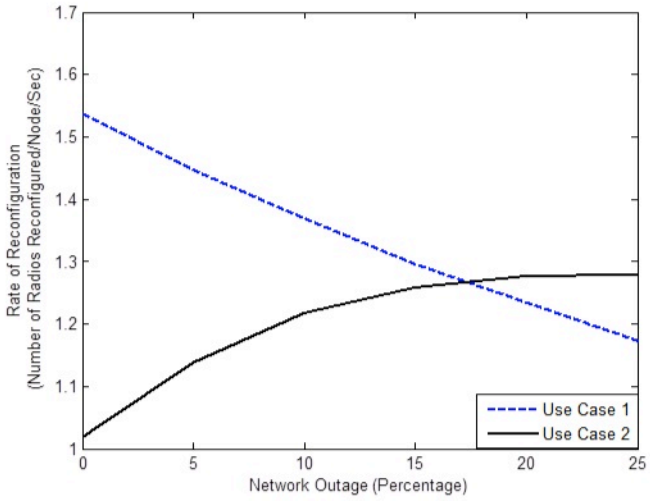

Figure 4.12: Reconfiguration rate for unbalanced network topology

Global allocation of resources in an integrated heterogeneous wireless environment that encompasses several RATs makes the resource allocation process more efficient by assigning each cUE in the system to the best APs/BSs. However, the gain in spectral efficiency comes at the expense of increased total power consumption. The tradeoff between spectral efficiency and power consumption largely depends on the nature of the heterogeneous network deployment assumptions. In our study, we showed the following trends:

- For a balanced deployment scenario, the gain in spectral efficiency for use case 2 compared to use case 1 is not very significant. The highest gain (75%) occurs when the network outage percentage is the highest (25%). For this network outage setting, users for both use case 1 and use case 2 experience significant number of reconfigurations

(reconfiguration rate of 1.25 and 1.45 respectively). As a result, the ratio of increase in power consumption for use case 2 compared to use case 1 is the least for this setting and lies in the range [32.2%, 614.3%] depending on varying hardware assumptions in terms of ASIC vs. FPGA circuitry. The lowest spectral efficiency gain (14.30%) between the two use cases occurs when the network outage percentage is the lowest (0%). In this case, the number of reconfigurations needed by use case 1 (0.2/second) is not as high as the one needed by use case 2 (0.7/second). As a result, the corresponding increase in power consumption for use case 2 compared to use case 1 is the highest, which lies in the range [129.8%, 916.8%]. So, the reconfiguration rate (or number of reconfigurations required by each user) mainly dictates the power consumption trends for balanced network deployment.

- For an unbalanced deployment scenario, the gain in spectral efficiency is significant. The highest gain (553.7%) occurs when the network outage percentage is the lowest (0%) and the lowest gain (314.3%) occurs when the network outage percentage is the highest (25%). For each network outage setting in the unbalanced scenario, resources of some Wi-Fi APs cannot be used by any user for use case 1, but these resources can be used by users for use case 2. So the actual power consumption not only depends on the reconfiguration rate, but also on the number of APs/BSs that are used. As a result, the highest increase in power consumption is experienced when the network outage percentage is 25%, which corresponds to the lowest gain in spectral efficiency. For this setting (25% network outage), the increase in power consumption lies in the range [88.6%, 614.9%] depending on hardware assumptions. The lowest increase in power

consumption occurs when the network outage percentage is 0%, which corresponds to the highest gain in spectral efficiency. For this setting (0% network outage), the increase in power consumption lies in the range [35.1%, 595.7%].

Based on these results, one can see that the hardware choices have a significant impact on the increase in power consumption. For the gain in spectral efficiency of up to 75%, the increase in power consumption can range from 32.2% to 916.8% for balanced deployment scenario, and for the gain in spectral efficiency of up to 553.7%, the increase in power consumption can range from 35.1% to 614.9% for the unbalanced deployment scenario depending on user device hardware assumptions. Depending on the level of reconfiguration that is required, for example, total number of reconfigurable radios (maximum of 2 in our simulated scenario), it might be possible to attain a tradeoff in terms of lower power consuming ASIC radios at the cost of decreased reconfigurable options. While low power reconfigurable fabrics with power consumption on the order of current ASIC technology are not available today, hybrid architectures that use both ASIC and FPGA components could provide a practical approach to reduce the power consumption of reconfigurable devices.

4.3 Optimization-based Algorithm Simulation Study

The focus of this next study is to quantify the tradeoffs achieved in terms of network efficiency measures of spectral efficiency, instantaneous and long term fairness, and energy consumption which pertains to any general network topology, network co-operation model, user device assumptions, user mobility patterns and network outage assumptions. As a result, for this study we limit the number of scenarios we consider for

each of these options compared to the previous study. In particular, for this study we limit our attention to the balanced network topology used in our earlier study presented in Figure 4.1. For network co-operation model, we consider both use case 1 and use 2 described in Chapter 3. However, we restrict the type of users to nomadic users. As a result, for use case 1, all users can connect only to their own carrier's cellular and Wi-Fi networks. For use case 2, all users can connect to any network (both carrier's cellular and Wi-Fi networks) in the topology. For both use cases, each user device is equipped with three reconfigurable radios that are implemented using FPGA fabric. All users move in the network topology using a random waypoint mobility model at a constant speed of 2 mph. Since we do not use a detailed channel model, the fluctuations in connectivity conditions are modeled by a random independent network outage of 5% for each AP/BS.

The GRC instructs each user device to configure (or reconfigure) its radios for use with the appropriate APs/BSs on a 1-second basis. For this study, we assume that there is no communication downtime during a reconfiguration handoff. However, the energy consumption cost during a reconfiguration handoff has two sub-components. The first sub-component, $E_{rec,a}$, represents the extra energy that is spent by radios in reconfiguring the hardware to transition to a new RAT. We assume an FPGA platform as our reconfigurable radio platform. The second sub-component, $E_{assoc,a}$, represents the extra energy that is spent associating with a new RAT. Values used for both $E_{rec,a}$ and $E_{assoc,a}$ in this study are presented in Table 3.3. We do not use the impact of reconfiguration experimental parameter in this study and identify its use as a part of future work.

The GRC uses a multi-attribute resource allocation algorithm to determine the

user device-to-AP/BS mappings and the rate assignment per mapping for each scheduling interval t. The attributes considered in this algorithm are system spectral efficiency, both instantaneous and long-term fairness in terms of data rate allocated to each user in the system, and battery lifetime of each user (or overall energy consumption) in the system. Since the achieved performance in terms of each of these attributes belongs to a different set of ranges, we normalize the performance achieved for each of these attributes on a [0,1] scale using a utility function approach. We describe the utility function for each attribute next using system parameters presented in Table 3.1.

(i) *Spectral Efficiency Utility Function*

The achievable system spectral efficiency for time interval [t, t+1], denoted γ^t, is computed as the ratio of the rate allocated to each user in the system at time t to the total spectrum used and is represented by (3.1). Since we assume that the amount of spectrum managed by each RAT is constant, the total spectrum, κ, used by our system remains constant. So, to maximize the achievable system spectral efficiency, the objective of any network optimization problem is to maximize the sum of the rates allocated to each user subject to total resource usage constraints. This optimization problem has been well studied as the max-sum rate (MSR) optimization problem. The idea behind the MSR optimization objective is to assign each resource block to the user that can make the best use of it. The drawback of the MSR optimization objective is that it is likely that a few users close to the BS, and hence having excellent channel conditions, will be allocated all the system resources. As a result, the MSR optimization objective cannot be used as the

only objective in any resource allocation problem. Fairness of resource distribution also has to be taken into account.

However, since the MSR optimization objective results in the highest achievable system spectral efficiency, it can be used as an upper bound in computing the spectral efficiency utility function. Let γ_{max}^{t} represent the achievable system spectral efficiency for time interval [t, t+1] obtained by solving the MSR optimization problem. Similarly, assuming each available resource block is allocated to some user, the minimum achievable system spectral efficiency results when each resource block is assigned to the user with worst connectivity conditions. Let γ_{min}^{t} represent this minimum achievable system spectral efficiency for time interval [t, t+1]. Then, γ_{min}^{t} can be used as a lower bound in computing the spectral efficiency utility function. The normalized system utility γ_{util}^{t} is then computed using (4.1). If the achievable system spectral efficiency equals γ_{max}^{t}, the spectral efficiency utility function corresponds to a value of 1, and if the achievable system spectral efficiency equals γ_{min}^{t}, the spectral efficiency utility function corresponds to a value of 0.

$$\gamma_{util}^{t} = \frac{\gamma^{t} - \gamma_{min}^{t}}{\gamma_{max}^{t} - \gamma_{min}^{t}} \tag{4.1}$$

(ii) *Long-Term Fairness Utility Function*

The long-term fairness metric is computed using Jain's Fairness Index as shown in (3.3). Since Jain's Fairness Index is already normalized in the range [0,1], we set long-term fairness utility function, ϕ_{util}, equal to the long-term fairness metric as presented in (4.2).

$$\phi_{util} = \phi = \frac{(\sum_{u \epsilon U} \sum_t r_u^t)^2}{|U| * \sum_{u \epsilon U} (\sum_t r_u^t)^2} \quad (4.2)$$

(iii) *Instantaneous Fairness Utility Function*

We assume support for real-time traffic in this study. So, the instantaneous fairness metric, θ_{RT}^t, is computed using (3.4). Moreover, since we use the ratio of blocked users to compute this metric, the obtained value for this performance metric is already normalized in the range [0,1] and hence the instantaneous fairness utility, θ_{util}^t, is equal to the instantaneous fairness metric, θ_{RT}^t, as shown in (4.3).

$$\theta_{util}^t = \theta_{RT}^t = 1 - \frac{|BU^t|}{|U|} \quad (4.3)$$

(iv) *Overall Energy Consumption Utility Function*

The energy consumption for user $u \in U$ during time interval [t, t+1], denoted as ω_u^t, is computed using (3.7). The goal of the overall energy consumption optimization is to minimize the overall energy consumed by each user in the system for each scheduling interval [t, t+1]. We use the same maximum and minimum achievable system spectral efficiency concepts adopted in the spectral efficiency utility function in computing the overall energy consumption utility function. Let ω_{max}^t represent the maximum achievable overall energy consumption and let ω_{min}^t represent the minimum achievable overall energy consumption for time interval [t, t+1]. Then the battery lifetime utility function for time t, denoted as ω_{util}^t, is computed using (4.4). If the achievable overall energy consumption equals ω_{min}^t, the battery lifetime utility function is 1 and if the achievable overall energy consumption equals ω_{max}^t, the battery lifetime utility function is 0.

$$\omega_{util}^{t} = 1 - \frac{\sum_{u \in U} \omega_{u}^{t} - \omega_{min}^{t}}{\omega_{max}^{t} - \omega_{min}^{t}} \tag{4.4}$$

The utility functions derived for each of the four metrics is used in computing the final achieved performance. Some of these utility functions such as γ_{util}^{t} and ω_{util}^{t} are incorporated directly into the resource allocation procedure, whereas alternative formulations are used in the resource allocation procedure to achieve instantaneous and long-term fairness.

4.3.1 Optimization-based Resource Allocation Algorithm

In this section, we present the resource allocation procedure that is used by the GRC to come up with user device-to-AP/BS mappings and the rate assignment per mapping. Since our heterogeneous wireless system supports both real-time and best effort traffic, the resource allocation problem follows a two-step approach. In the first step, an iterative admission control policy is implemented to satisfy minimum data rate requirements (for real-time traffic) of as many users in the system as possible. In the second step, the weighted spectral efficiency, long-term fairness, and overall energy consumption utility functions (related to best-effort traffic) are maximized, subject to minimum data rate requirements. Algorithm 4.1 describes the complete resource allocation procedure that is used during each time step t.

Each step (Step 1 and 2) in the algorithm uses a mixed integer linear program (MILP) presented by (4.6) and (4.8) respectively. The objective of both MILPs is to determine x_{ua}^{t} (the assignment variable) and r_{ua}^{t} (the rates allocated to each radio of each user). The spectral efficiency, long-term fairness and overall energy consumption utility

functions are then computed using these x_{ua}^t and r_{ua}^t variables using (4.1), (4.2) and (4.4) respectively. Note that the overall energy consumption function presented in (4.4) depends on (3.7) which uses an additional variable η_{ua}^t, the maximum amount of data (in bits) that can be transferred by radio $a \in A$ of user $u \in U$ during the scheduling interval t. Since the GRC scheduler operates on a 1 second basis, η_{ua}^t equals r_{ua}^t in our study.

Algorithm 4.1 Multi-Attribute Resource Allocation

Step 0: Initialization

1. $if\ t == 1$
2. $\quad A_u^t \leftarrow 1 \quad \forall u \in U$
3. $\quad x_{ua}^{t-1} \leftarrow 0 \ \ \forall u \in U, \forall a \in A$
4. $\quad \rho \leftarrow 0.10$
5. $end\ if$

Step 1: Admission Control

6. $BU^t \leftarrow \emptyset, z \leftarrow infeasible,$
7. $while\ z$ is infeasible
8. $\quad select\ any\ u \in U, u \notin BU^t$
   $\quad z \leftarrow solve\ P^*\ \ using$ (4.6)
9. $\quad if\ z$ is infeasible
10. $\quad\quad r_{u,max}^t = \sum_{a \in A} r_{ua,max}^t / \mathrm{T}_u^t \quad \forall u \in U, u \notin BU^t$
11. $\quad\quad u_{drop} \leftarrow u \in \arg\min \{ r_{u,max}^t \}$
12. $\quad\quad BU^t \leftarrow BU^t \cup \{u_{drop}\}$
13. $\quad end\ if$
14. $end\ while$

Step 2: Multiple-Attribute Optimization

15. $solve\ MA^*\ \ using$ (4.8)
16. $A_u^{t+1} = (1-\rho)A_u^t + \rho r_u^t$

The goal of the admission control procedure, described by Step 1 in the algorithm, is to determine when a user is blocked and maximize the instantaneous fairness utility metric presented in (4.3) by minimizing the number of blocked users. The admission

control procedure first initializes the list of blocked users at time t (BU^t) to null and sets z, the variable that determines the feasibility of satisfying real-time traffic demands of each user, to be infeasible. Next, it recursively solves optimization problem P^*, using (4.6), in an effort to find a feasible solution that tries to satisfy the real-time traffic demand of each user using constraint (4.6b). Note that in formulating P^*, $r_{ua,norm}^t$ is used rather than r_{ua}^t in constraints (4.6c)-(4.6f) to avoid non-linear problem formulations. The relationship between r_{ua}^t and $r_{ua,norm}^t$ is described by (4.5). This relation removes the dependence of r_u^t on two variables, x_{ua}^t and r_{ua}^t as presented in (3.2). Now, r_u^t only depends on r_{ua}^t, as presented by (4.6a), as constraint (4.6d) makes sure that $r_{ua,norm}^t$ (and consequently r_{ua}^t) is greater than zero only if x_{ua}^t equals one. After solving one iteration of P^*, the admission control procedure checks whether a feasible solution is produced. If the solution to P^* is infeasible, the user with the worst achievable data rate to demand ratio is dropped and this user is added to the list of blocked users (BU^t) that are assigned no resource blocks (or are assigned rate 0 as described by constraint (4.6e)). The admission control procedure keeps solving P^* and dropping the user with worst achievable data rate to demand ratio until all users that are to be allocated resources ($u \in U, u \notin BU^t$) can achieve a data rate of at least T_u^t bits/s. This mechanism enables the admission control procedure to block as few users as feasible. Once a feasible solution is produced for P^*, the resource allocation procedure moves to Step 2 of the algorithm.

$$r_{ua,norm}^t = \frac{r_{ua}^t}{r_{ua,max}^t} \tag{4.5}$$

$$P^*:\ max\ r_u^t = \sum_{a \in A} r_{ua}^t \qquad (4.6a)$$

$$s.t.\quad r_u^t \geq \mathrm{T}_u^t \qquad \forall u \in U, u \notin BU^t \qquad (4.6b)$$

$$\sum_{u \in U} r_{ua,norm}^t \leq 1 \qquad \forall a \in A \qquad (4.6c)$$

$$r_{ua,norm}^t \leq x_{ua}^t \qquad \forall u \in U, u \notin BU^t, \forall a \in A \qquad (4.6d)$$

$$r_{ua,norm}^t = 0 \qquad \forall u \in BU^t, \forall a \in A \qquad (4.6e)$$

$$r_{ua,norm}^t \geq 0 \qquad \forall u \in U, u \notin BU^t, \forall a \in A \qquad (4.6f)$$

$$\sum_{a \in A} x_{ua}^t \leq m_u \qquad \forall u \in U, u \notin BU^t \qquad (4.6g)$$

$$x_{ua}^t \in \{0,1\} \qquad \forall u \in U, u \notin BU^t, \forall a \in A \qquad (4.6h)$$

The final step (Step 2) in the algorithm comes up with user device-to-AP/BS mappings and the rate assignment per mapping based on an optimization function, MA^*, described by (4.8), that optimizes the weighted spectral efficiency, long-term fairness and energy consumption utility functions subject to the minimum data rate requirements confirmed by the admission control procedure. The utility functions described in (4.1) and (4.4) are used in MA^* to maximize system spectral efficiency and minimize overall energy consumption, respectively. For long-term fairness, the utility function described by Jain's fairness index in (4.2) is non-linear and hard to solve for a large-scale heterogeneous wireless system. As a result, an alternative formulation that uses the ratio of instantaneous to average data rate described in (4.7) is used to maximize long-term fairness utility[2]. It has been shown that allowing the user with maximum achievable ratio of instantaneous to average data rate to transmit during each time step results in

[2] Note that ϕ'^{t}_{util} presented in (4.7) is only used in solving MA^*. ϕ_{util} representing Jain's fairness index in (4.2) is still used in computing long-term fairness utility.

maximizing fairness over long time scales [97]. Again, the maximum and minimum achievable ratios of instantaneous to average data rate are used in (4.7) to scale the long-term fairness utility between 0 and 1. The algorithm initializes the average data rate of each user $u \in U$, denoted as A_u^t, to 1 during the first time step as described in the initialization step in Algorithm 4.1. After solving the MA^* optimization problem, the algorithm updates the average data rate of each user over a time window that is dictated by the scalar ρ. The value of this scalar is commonly set between 0.05 and 0.10 [98]. We set $\rho = 0.10$ in our work as noted in the initialization step in Algorithm 4.1.

$$\phi_{util}'^{t} = \frac{\sum_{u \in U} \left[\frac{r_u^t}{A_u^t} - \left(\frac{r_u^t}{A_u^t} \right)_{min} \right]}{\sum_{u \in U} \left[\left(\frac{r_u^t}{A_u^t} \right)_{max} - \left(\frac{r_u^t}{A_u^t} \right)_{min} \right]} \tag{4.7}$$

$$MA^*: \max \ (\alpha * \gamma_{util}^{\ t}) + \ (\beta * \phi_{util}'^{t}) + \ (\tau * \ \omega_{util}^t \ ) \tag{4.8a}$$

$$s.t. \quad r_u^t \ \geq \mathrm{T}_u^t \qquad \forall u \in U, u \notin BU^t \tag{4.8b}$$

$$\sum_{u \in U} r_{ua,norm}^t \leq 1 \quad \forall a \in A \tag{4.8c}$$

$$r_{ua,norm}^t \ \leq x_{ua}^t \qquad \forall u \in U, u \notin BU^t, \forall a \in A \tag{4.8d}$$

$$r_{ua,norm}^t \ = 0 \qquad \forall u \in BU^t, \forall a \in A \tag{4.8e}$$

$$r_{ua,norm}^t \ \geq 0 \qquad \forall u \in U, u \notin BU^t, \ \forall a \in A \tag{4.8f}$$

$$\sum_{a \in A} x_{ua}^t \leq m_u \qquad \forall u \in U, u \notin BU^t \tag{4.8g}$$

$$x_{ua}^t \in \{0,1\} \qquad \forall u \in U, u \notin BU^t, \forall a \in A \tag{4.8h}$$

Note that as stated earlier, we assume a user device can use multiple radios concurrently. The maximum number of radios that a user device can concurrently use is limited by m_u variable presented in (4.6g) and (4.8g). In our problem formulation, we

assume $m_u(=3)$ to be the same for each user. There might be cases where the value of m_u can vary for different users. For example, if a user device does not have enough energy to support more than one physical link (i.e. the device is operating at a low battery level), then a policy-based addition can be included in the algorithm that limits such a user to use only one of its radios. These policy-based decisions represent a possible extension to our current model.

The scalars $\alpha, \beta \; and \; \tau$ provide the relative importance of each optimization attribute in $MA^*$and act as 'control knobs' that allow network operators to achieve the desired performance objectives. The values for these scalars are obtained through AHP [73]. AHP is a decision analysis technique to determine weights of different utility attributes from decision stakeholders through pairwise comparisons and ratings. Using AHP, we interviewed two experts from the cellular industry to perform pairwise comparisons between our utility attributes[3]. After determining which attribute is more important, the more important attribute receives a score from 1-9, with 1 indicating that the two attributes are equally important. These pairwise comparisons are placed in matrix **A**, with $a_{ji} = 1/a_{ij}$, where each row and column represents a specific attribute. Using the following equation: $\mathbf{Aw} = \lambda_{max}\mathbf{w}$, and solving for λ_{max}, the principal eigenvalue of **A**, and **w**, the principal right eigenvector of **A**, we can normalize the entries of **w** by dividing by their sum and recover the weighted values for our utility function.

We asked each expert to compare the relative importance of battery life, fairness, and efficiency [101]. The results of the interview are placed in a comparison matrix, from

[3] While in this work we only examine only two viewpoints, we also note that group decision-*making* and viewpoint aggregation has been studied in [99, 100]

which the principal eigenvector is calculated. The results from this calculation and resulting weight values are shown in Table 4.1. From Table 4.1, we note that results of AHP show that both experts had relatively similar weight preferences. Consequently, we use results derived from Expert 1's responses in the remainder of our work.

Table 4.1: AHP matrices derived from expert interviews

Expert 1				
	Battery Life	Long-Term Fairness	Spectral Efficiency	Weights
Battery Life (BL)	1.0	5.0	0.333	0.279
Long-Term Fairness (LTF)	0.2	1.0	0.143	0.072
Spectral Efficiency (SE)	3.0	7.0	1.0	0.649
Expert 2				
	Battery Life	Long-Term Fairness	Spectral Efficiency	Weights
Battery Life	1.0	5.0	0.500	0.333
Long-Term Fairness	0.2	1.0	0.143	0.075
Spectral Efficiency	2.0	7.0	1.0	0.592

4.3.2 Optimization-based Algorithm Results and Analysis

Each simulation is run in MATLAB for 10,000 seconds. We first present results for when wireless data networks only support best-effort traffic. For this case, there is no minimum data rate requirement for any user. In other words, $T_u^t = 0$ for all users in the system. Since $T_u^t = 0$, the admission control procedure does not block any user for any scheduling time step and is not needed. As a result, the instantaneous fairness utility metric is not computed for this case. The overall utility function only depends on the spectral efficiency utility (γ_{util}), long-term fairness utility (ϕ_{util}) and energy consumption utility (ω_{util}), averaged over the entire simulation run, and is calculated using (4.9) where α = 0.649, β = 0.072, and τ = 0.279. We provide the overall utility

results with each of the three utility components for use case 1 and use case 2 in Figures 4.13 and 4.14 respectively. Optimization problems presented in (4.6) and (4.8), which are parts of the proposed algorithm, are solved using AMPL modeling language and CPLEX optimization solver [102-103].

$$Overall_{util,BE} = (\alpha * \gamma_{util}) + (\beta * \phi_{util}) + (\tau * \omega_{util}) \quad (4.9)$$

In addition to the utility results for our multi-attribute resource allocation algorithm, we provide results for four commonly used scheduling algorithms for wireless data networks: (i) min power (ii) max-sum rate (iii) proportional fairness and (iv) max-min fairness. Note that the first three algorithms reduce to our MA^*optimization if we set (i) $\alpha = 0$, $\beta = 0$, $\tau = 1$ (ii) $\alpha = 1$, $\beta = 0$, $\tau = 0$ and (iii) $\alpha = 0$, $\beta = 1$, $\tau = 0$ respectively in (4.8a). The max-min fairness results are obtained using the progressive filling algorithm [104]. Furthermore, the max-sum rate algorithm always achieves the highest system spectral efficiency and as a result its $\gamma_{util} = 1$ for both use cases. However, because of more connectivity options for use case 2, the average spectral efficiency for use case 2 is 4.35 bits/s/Hz compared to 3.52 bits/s/Hz for use case 1. Similar to the max-sum rate algorithm, the min power algorithm always produces the minimum possible energy consumption and therefore its $\omega_{util} = 1$ for both use cases. But the average energy consumption per user is 9600 Joules for use case 1 compared to 10400 Joules for use case 2. All other algorithms compute their spectral efficiency utility relative to max-sum rate algorithm's spectral efficiency utility as described by (4.1) and their energy consumption utility relative to min power algorithm's energy consumption utility as described by (4.4).

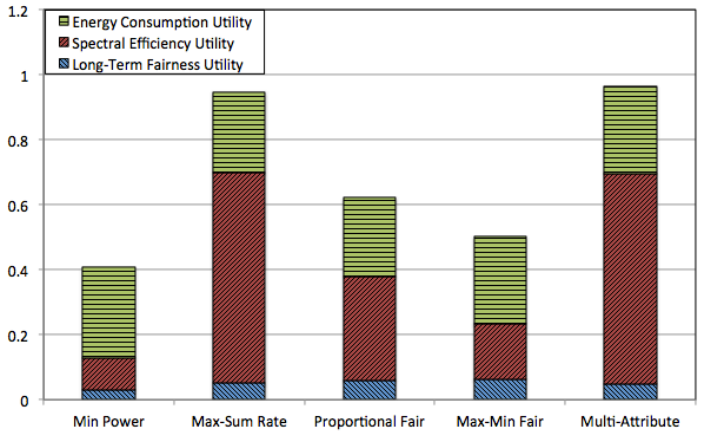

Figure 4.13: Overall utility for use case 1, $T_u^t = 0$

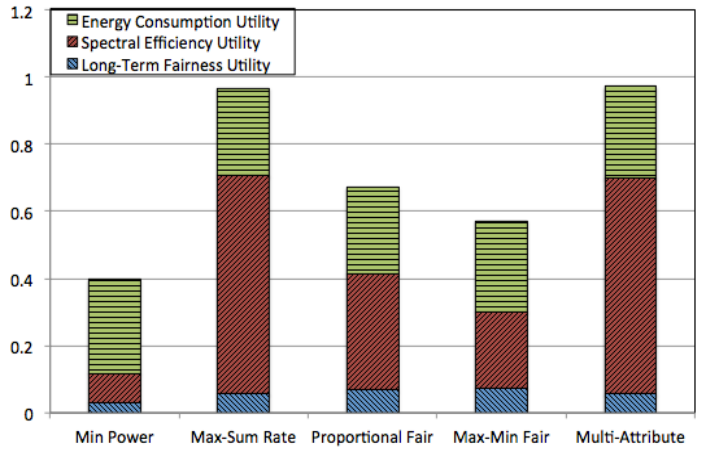

Figure 4.14: Overall utility for use case 2, $T_u^t = 0$

The overall utility of our multi-attribute resource allocation algorithm is very similar to the overall utility of max-sum rate algorithm for both use case 1 (0.967 compared to 0.948) and use case 2 (0.971 compared to 0.967) as seen from Figures 4.13 and 4.14 respectively. Since the spectral efficiency utility is given the highest weight in our overall utility function, this result follows expectations. In comparison to the max-sum rate algorithm, our algorithm improves the energy consumption utility (0.269 compared to 0.247 for use case 1 and 0.271 compared to 0.260 for use case 2) at the cost of a slight degradation in spectral efficiency utility (0.648 compared to 0.649 for use case

1 and 0.644 compared to 0.649 for use case 2). The long-term fairness utility is almost the same for our algorithm and the max-sum rate algorithm for both use case 1 (approximately 0.050) and use case 2 (approximately 0.056). All other algorithms (min power, proportional fairness, max-min fairness) sacrifice spectral efficiency in trying to achieve other objectives, as seen in Figures 4.13 and 4.14, and as a result their overall utility is much lower than the one obtained by our algorithm.

We now consider the case of next-generation heterogeneous wireless networks that are expected to support both real-time and best-effort traffic. In this case, the overall utility function depends on utility attributes that apply to real-time traffic and the attributes that apply to best-effort traffic. We equally weigh the utilities of both traffic types to compute the overall utility function. The best-effort traffic utility, denoted $Overall_{util,BE}$, depends on spectral efficiency, long-term fairness and energy consumption utilities as presented in (4.9). The real-time traffic depends on instantaneous fairness utility averaged over the entire simulation run, denoted θ_{util}, and is calculated using (4.3). Hence, the overall utility function for next-generation heterogeneous wireless networks is computed using (4.10).

$$Overall_{util,BE+RT} = \frac{1}{2} * Overall_{util,BE} + \frac{1}{2} * \theta_{util} \tag{4.10}$$

For the next-generation heterogeneous wireless networks, we present results for both use case 1 and use case 2 using Figures 4.15 and 4.16, respectively, where the minimum data rate requirement of each user to support real-time traffic is T_u^t = 512 kbps. The overall utility of our algorithm for both use cases is significantly higher than any

other algorithm. For both use cases, the overall energy consumption utility and long-term fairness utility of all algorithms are similar. But the difference in overall utility is obtained due to instantaneous fairness and spectral efficiency utilities. For use case 1, in terms of overall utility performance, our algorithm outperforms the next closest algorithm, max-sum rate, by 56.7% (0.818 compared to 0.522). The spectral efficiency utility of our algorithm for best-effort traffic decreases compared to max-sum rate algorithm (0.224 compared to 0.325). But this happens as a result of satisfying more real-time traffic users. The instantaneous fairness utility of our algorithm is significantly higher than that of max-sum rate algorithm (0.437 compared to 0.048). For use case 2, our algorithm outperforms the next closest algorithm, max-min fairness, in terms of overall utility by 24.0% (0.975 compared to 0.786). The instantaneous fairness utility of both algorithms is 0.5. But the spectral efficiency utility of our algorithm is significantly higher compared to max-min fairness algorithm's spectral efficiency utility (0.310 compared to 0.115). This shows that for future heterogeneous wireless systems supporting both real-time and best-effort traffic, our algorithm always obtains the best of both worlds by applying the right trade-offs in terms of achieved spectral efficiency and instantaneous fairness.

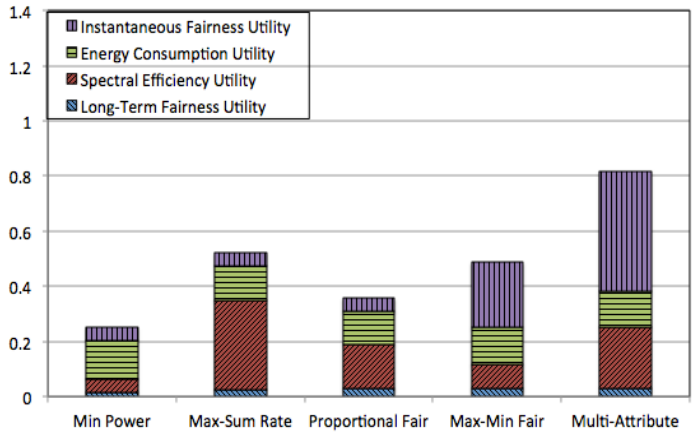

Figure 4.15: Overall utility for use case 1, T_u^t = 512K

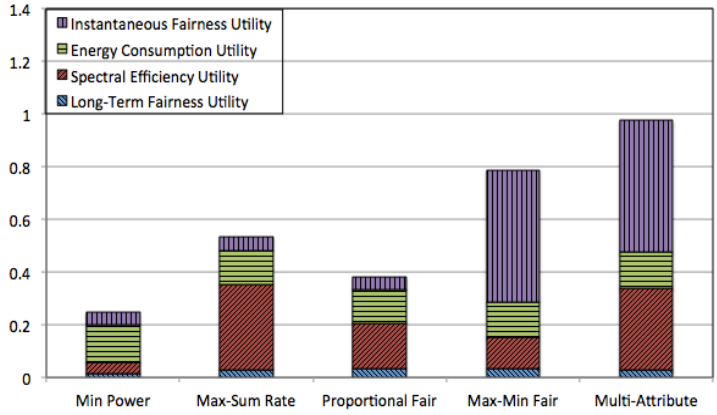

Figure 4.16: Overall utility for use case 2, T_u^t = 512K

We finally present results for both use case 1 and use case 2 for future heterogeneous wireless systems for different levels of minimum data rate requirements using Figures 4.17 and 4.18 respectively. Note that we still assume that each user has identical requirements T_u^t, but we study the effects of varying values of T_u^t. In both use cases for all different levels of T_u^t, our algorithm outperforms any other algorithm. None of the other algorithms is suited to support both best-effort and real-time traffic. While max-sum rate and proportional fairness algorithms are well suited for achieving good spectral efficiency for best-effort traffic, they do not provide acceptable levels of

instantaneous fairness. On the other hand, the max-min fairness algorithm provides good instantaneous fairness, but its spectral efficiency suffers significantly. Our algorithm achieves a balance in both instantaneous fairness and spectral efficiency utilities. Apart from this, there are two additional observations of interest in Figures 4.17 and 4.18. First, while most traditional algorithms provide constant overall utility levels and then possibly experience sudden drops in performance (for example, max-min fairness algorithm for use case 2), our algorithm degrades gradually as the available resources cannot satisfy the demands. Second, since use case 2 represents more connectivity options for each user, the resulting overall utility of our algorithm is considerably higher (by up to 39.4%) compared to use case 1 for higher levels of T_u^t ($T_u^t \geq 512$ kbps). So increasing the amount of connectivity options (possibly through peering agreements among several network service providers) has significant performance benefits.

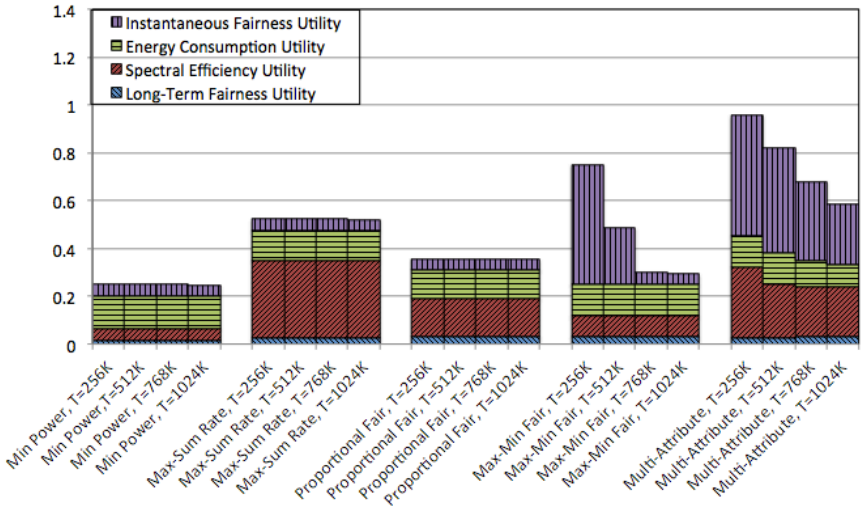

Figure 4.17: Overall utility for use case 1, variable T_u^t

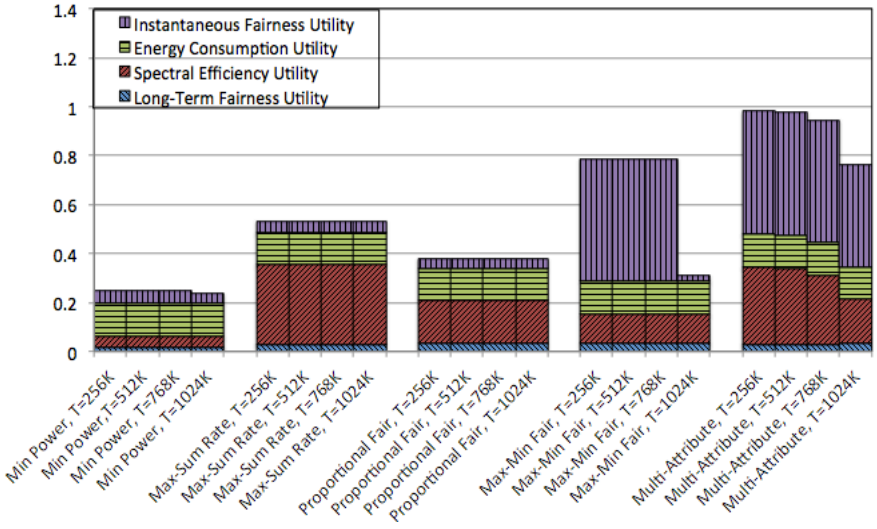

Figure 4.18: Overall utility for use case 2, variable T_u^t

The main conclusions of our optimization-based study can be summarized as follows:

- The traditional algorithms achieve good performance in terms of one or two attributes, but they suffer in terms of other attributes. The max-sum rate algorithm achieves good spectral efficiency but suffers in terms of instantaneous fairness. The proportional fairness algorithm achieves good spectral efficiency and long-term fairness, but suffers in terms of instantaneous fairness. The max-min fairness algorithm achieves good long-term fairness, but suffers in terms of spectral efficiency. The min power algorithm achieves good energy consumption, but suffers in terms of spectral efficiency and instantaneous fairness.
- By following a two-step resource allocation procedure, depending on the situation, our algorithm improves the overall system performance by achieving the right trade-offs in terms of system spectral efficiency and energy consumption (for best-effort traffic) or by achieving the best trade-offs in terms of system spectral efficiency and

instantaneous fairness (for real-time traffic).

- Through MATLAB/CPLEX based simulations, we showed an increase in overall utility of up to 56.7% for our algorithm compared to the next best algorithm.

CHAPTER FIVE

PRACTICAL IMPLEMENTATION ISSUES

For the second phase of our research study, we use a detailed protocol level simulator, ns-2, to study the resource management problem in a hetnet system. To the best of our knowledge, no prior work has considered the impacts of RAT-specific implementation details while assessing the benefits of a hetnet system. Most of the prior work related to the hetnet system has focused on studies (analytical or simulation) based on simplified network assumptions (similar to our work in the first phase). In this phase, we explore the management overhead required for a hetnet system where user device-to-AP/BS associations are controlled by a centralized GRC. Using our four network efficiency measures of spectral efficiency, instantaneous fairness, long-term fairness, and overall energy consumption, we show the performance benefits of a hetnet system where user device-to-AP/BS associations are controlled by a centralized GRC compared to a distributed solution. The performance benefits are analyzed for two greedy sort-based algorithms implemented at the GRC that try to maximize system spectral efficiency and instantaneous fairness respectively. For the investigations conducted in the first phase of our research, the GRC scheduling algorithm accounts for real-time traffic in addition to best-effort traffic as it is based on high-level simulation model. For this phase, we focus only on best-effort traffic, as the main goal of this work is to study the impact of RAT-specific implementation issues and centralized control overhead on achieved network performance. We also identify technical challenges associated with periodic re-

associations proposed by our hetnet solution and provide possible alternatives to remedy the challenges.

5.1 Problem Assumptions

The hetnet system that we consider for this phase of work consists of Wi-Fi and WiMAX RATs, GRC, and multi-modal user devices. We make the following assumptions related to these entities present in our hetnet system:

(i) The MAC protocol implemented at Wi-Fi APs achieve max-min fairness on long time scales; the MAC protocol implemented at WiMAX BSs can achieve max-min fairness or proportional fairness on both short and long time-scales: The scheduler implemented at the MAC layer of each AP/BS deployed in practice has a predefined scheduling objective. Wi-Fi MAC implements a standardized DCF solution that employs CSMA/CA with binary exponential backoff algorithm. Moreover, the Wi-Fi AP implements a FIFO queuing system where each arriving packet is served in order. It has been shown that the DCF MAC and FIFO queuing mechanism implemented in a Wi-Fi system leads to equal throughput for all associated user devices on a long time scale [80-81]. We implement this standard DCF MAC and FIFO queuing mechanism for the Wi-Fi RAT used in our study and hence users connected to Wi-Fi RATs achieve 'local' long-term max-min fairness in our system. The WiMAX standard leaves the scheduler implementation at the MAC layer up to the BS equipment/service provider. Max-min fairness and proportional fairness resource allocation schemes have been studied extensively in literature as a means of sharing resources fairly among all connected users and proposed as the likely objectives for a scheduler implemented for 4G RATs. We use

a deficit weighted round robin (DWRR) scheduler at the WiMAX BS. By tuning the weight associated to the data packet queues for each connected user to an appropriate value, the WiMAX BS in our system can achieve either max-min fairness or proportional fairness objectives for both short and long time-scales. The GRC uses the scheduling objective (max-min fairness or proportional fairness) information for both WiMAX BSs and Wi-Fi APs while computing user device-to-AP/BS association decisions each scheduling interval. Note that for this solution, GRC does not have to relay the data rate per association mapping information to the Wi-Fi APs/WiMAX BSs.

(ii) Media Independent Handover (MIH) function has been implemented at layer 2.5 of the OSI stack at each AP/BS, user device and GRC for supporting messaging framework required for a centralized solution: The information related to the message exchanges required for this IEEE 802.21-based centralized solution is presented in detail in the next section (Chapter 5.2). Both WiMAX and Wi-Fi RATs in our system use an adaptive MCS. The signal strength at which the management/data packets are received on a user device (which is based on the distance of the user device from the corresponding BS/AP) dictate the MCS used by the radios on the device to connect to the corresponding BS/AP. The fast feedback channel quality indicator (CQICH) data burst in the uplink sub-frame has been implemented for the WiMAX MAC in ns-2 to relay the MCS update information to the WiMAX BS. An ACK piggyback mechanism has been implemented for the Wi-Fi MAC in ns-2 to relay the MCS update information to the Wi-Fi AP. The MCS dictates the maximum achievable data rate ($r^{t}_{ua,max}$ parameter presented in Table 3.1) for each radio on each user device. The details on maximum achievable data rates for

both WiMAX and Wi-Fi RATs used in our ns-2 study are presented in Tables C1-C2 in Appendix C. All the data packets are transmitted using the adaptive MCS and the associated data rate. However, all control messages related to each MAC and also the MIH messages are transmitted using the most robust MCS (BPSK 1/2) for both WiMAX and Wi-Fi.

(iii) Each user device is equipped with two static radios (ASIC-based hardware) but can only use one radio for an active data connection at a time: We consider the integral association scenario used in practice today, where only one radio on a multi-modal device can be used at any given time. Extensions in the networking stack are required to support multi-radio multi-flow capability assumed by the fractional association scenario. The optimization problem (in terms of overall system throughput or fairness) of coming up with an integral association in a heterogeneous wireless network environment is shown to be NP-hard [105-106]. Therefore, we limit the resource allocation studies for this work to heuristic algorithms.

(iv) The GRC operates on a five-second scheduling interval: The GRC computes the user device-to-AP/BS association decisions each scheduling interval by considering the independent scheduling objective (max-min fairness or proportional fairness) for both WiMAX and Wi-Fi RATs. While the independently implemented DWRR scheduler at WiMAX MAC converges to a max-min fairness or proportional fairness objective on short time-scales (milliseconds), the DCF-based Wi-Fi MAC converges to max-min fairness objective on larger time-scales (seconds). To allow the Wi-Fi MAC to converge to the max-min fairness solution and also to account for issues such as result propagation

delay, the GRC performs global-level optimizations (re-association computations) every five seconds.

(v) Infinitely backlogged downlink TCP traffic: We consider data traffic flow in the downlink direction (from BS/AP to user device). We only study best-effort traffic that is transmitted at a constant bit rate (CBR) over TCP transport layer. The traffic is sent at a rate higher than what could be supported by any RAT. So, the data connection queues supporting the TCP traffic for each user device at the AP/BS are almost always fully backlogged. However, when a TCP timeout occurs because of handovers or collisions, the TCP protocol performs the Additive Increase Multiplicative Decrease (AIMD) congestion control mechanism. As a result, at certain times the data connection queues of a few connected users at each AP/BS might not be full.

5.2 Extended System Model

The interaction between GRC, AP/BS of each RAT and each user device is handled via Media Independent Handover Function (MIHF) defined by the IEEE 802.21 standard. The mobility package provided by NIST [107] is used to implement the MIHF functionality in ns-2. The MIHF functionality is implemented at Layer 2.5 of the OSI stack as shown in Figure 5.1. The MIHF defines three different services: Media Independent Event Service (MIES), Media Independent Command Service (MICS) and Media Independent Information Service (MIIS). MIES provides events triggered by changes in the link characteristic and status. MICS provides the user devices necessary commands to manage and control the link behavior of each radio to accomplish handover functions. MIIS provides information about the neighboring networks and their

capabilities. We make use of MIES and MICS functionalities to manage the link-layer (Layer 2 of the OSI stack) network re-associations in our proposed hetnet solution. The messages related to each of these two services that are used in our study are summarized in Tables 5.1 and 5.2 respectively.

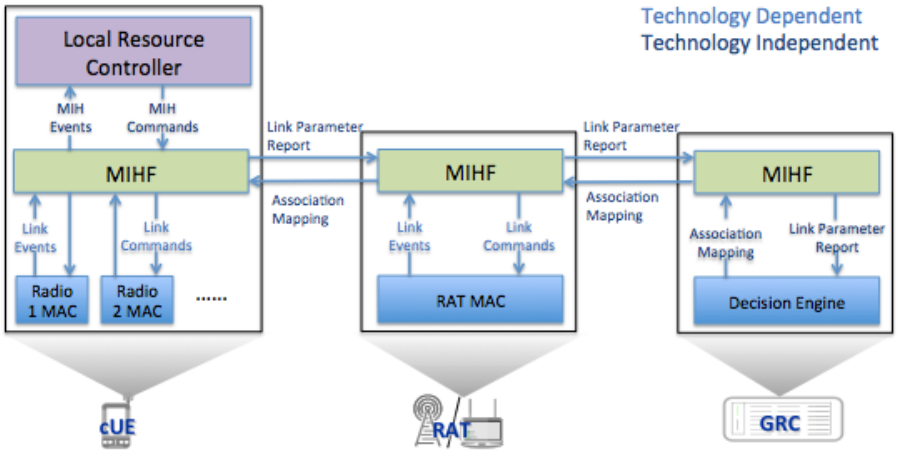

Figure 5.1: MIHF implementation in ns-2

From the events and commands presented in Tables 5.1 and 5.2, only association mapping event service and link parameter report command service generate actual packet overhead as messages related to these events/services are exchanged between two different entities (cUE, GRC). The association mapping and link parameter report messages are technology independent and are sent over the radio that is active at the corresponding cUE at the time these triggers are generated. All other messages are locally generated and aid cUE in managing its local interfaces. However, as noted in Table 5.2, technology dependent association/scan procedures, which follow a link connect/link scan trigger, might generate technology-specific message overhead.

Table 5.1: Media independent event services

Event Trigger	Trigger Generating Entity	Description
Link Up	cUE	Generated when any radio in cUE establishes link-layer connectivity with a BS/AP.
Link Down	cUE	Generated when any radio in cUE loses connectivity with a BS/AP. For Wi-Fi radio, this happens when 10 consecutive beacon packets (sent every 100 ms) are lost. For WiMAX radio, this happens when 120 consecutive DL-MAP/UL-MAP messages (sent every 5 ms) are lost.
Link Going Down	cUE	Generated when any radio in cUE receives a packet whose signal strength is lower than LGD_Threshold (= 1.1) * Rx Threshold.
Link Detected	cUE	Generated by any radio in cUE that receives synchronization messages (beacon for Wi-Fi and DL-MP for WiMAX) from AP/BS to which it is not currently connected.
Association Mapping	GRC	Generated by the GRC after computing periodic re-associations based on the decision engine (resource allocation procedure). This message is only sent to cUEs whose current network association must change.

Table 5.2: Media independent command services

Command Trigger	Trigger Generating Entity	Description
Link Connect	cUE	Generated when the MIHF in cUE wants one of its radios to establish a data connection with new BS/AP. Once this trigger is received by the corresponding radio, technology dependent association procedure follows.
Link Scan	cUE	Generated when the MIHF in cUE wants one of its radios to scan for BSs/Aps. Once this trigger is received by the corresponding radio, technology dependent scanning procedure follows.
Link Parameter Report	cUE	Generated periodically by cUE to send current link parameter status information (such as achievable MCS) related to all of its radios to the GRC.

The procedural flow of a re-association process is shown in Figure 5.2. During a re-association process, after a radio establishes link-layer (Layer 2) connectivity, IP connectivity (Layer 3) has to be established before a data flow can be directed to the new connection. We use the neighbor discovery protocol for obtaining an IPv6 address to establish Layer 3 connectivity [108]. Upon establishing Layer 2 connectivity, a *router solicitation* message is broadcasted by the radio. We assume that the neighbor discovery protocol functionality has been implemented at each AP/BS. When the AP/BS receives the router solicitation message, it sends a *router advertisement* broadcast packet in response. Upon receiving the router advertisement packet, the cUE uses the prefix information of the router advertisement packet to determine its new IP address. Moreover, to account for cases where a router solicitation/advertisement message is lost, the AP/BS broadcasts the router advertisement packet periodically so that a radio waiting for a new IP address due to packet loss can obtain the required address. Once IP connectivity is established, any flow in the uplink direction can start using the new radio connection. For the flow in the downlink direction, the other end-point of the flow has to be informed of the new IP address data packets need to be sent on. A *flow redirect request* message is sent by the cUE to the other end-point to accomplish this task. Upon receiving the flow redirect request message, the other end-point starts sending packets for the corresponding cUE to the new destination IP address. Moreover, the other end-point sends an ACK packet (of negligible size) back to the cUE to inform the cUE of the reception of flow redirect request message. The flow redirect request message is retransmitted by the cUE until an ACK packet is received from the other end-point. From

an overhead perspective, we consider router solicitation, router advertisement responses to the router solicitation messages, and flow redirect request messages as overhead messages required by the hetnet solution. The periodic retransmission of router advertisement packets is defined by the neighbor discovery protocol standard and would apply to any network using this IP address discovery method. Hence, we do not use all router advertisement messages as overhead messages.

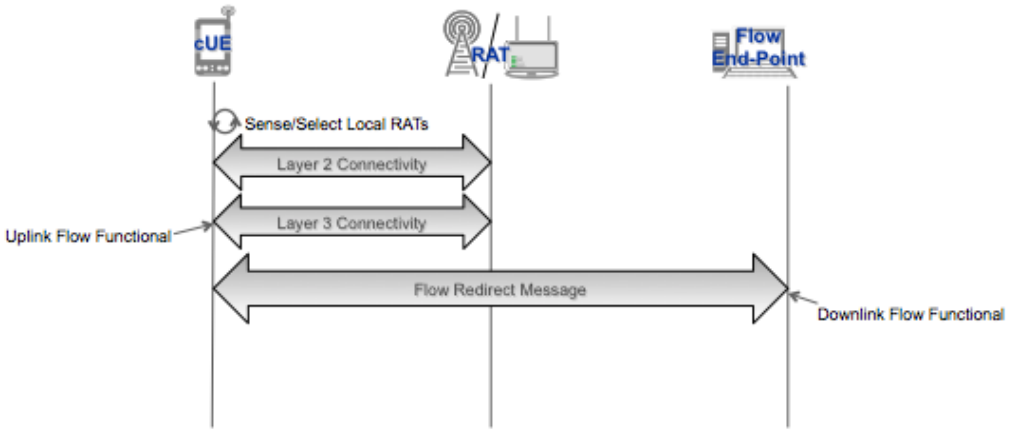

Figure 5.2: Procedural flow of a re-association process

The information related to Layer 2 and Layer 3 overhead messages used in our study is summarized in Table 5.3.

Table 5.3: Overhead messages in ns-2

Message	OSI Layer	Message Contents	Packet Size (Bytes)
Association Mapping	2	New Link Type, New Point-of-Attachment (PoA)	42
Link Parameter Report	2	Current Link Type, Current PoA, [Link Type, PoA, MCS] for all detected RATs OR current location	60
Router Solicitation	3	Route Request	48
Router Advertisement	3	Route Reply	96
Flow Redirect Request	3	Redirect IP Address	48

5.3 Greedy Sort-based Algorithm Simulation Study

The focus of this study is to show the performance benefits of a hetnet system where user device-to-AP/BS associations are controlled by a centralized GRC compared to a distributed solution and to identify technical challenges associated with a centralized scheme that performs periodic re-associations including the quantification of management overhead required by such a system. We consider a 2 * 2 km^2 grid where six Wi-Fi APs spread evenly throughout the topology and two WiMAX BSs located near the center of the grid are available to users for data connectivity. The simulation topology we consider for this study is presented in Figure 5.3. The coverage range of Wi-Fi AP is 0.15 km and the coverage range of WiMAX BS is 1 km. Note that the two WiMAX BSs have overlapping coverage area. However, both BSs operate on different frequency bands and thus avoid the interference co-ordination problem. The network topology in our simulation is similar to the balanced network topology used in our earlier studies presented in Figure

4.2. However, we do not consider any network co-operation model for this study. We assume each user device, based on its location, can connect to any available AP/BS in the area (similar to our use case 2 network co-operation model from earlier studies). There are 100 user devices in the 2 * 2 km^2 simulation topology. Each user device is equipped with a static Wi-Fi and WiMAX radio that are implemented using low energy consuming ASIC-based hardware. Each user receives a CBR data flow over TCP transport layer from the sink node. The relevant simulation parameters related to the Wi-Fi RAT, WiMAX RAT and the data flow are presented in Appendix D.

All users move in the network topology using one of three user movement patterns: (i) Linear movement pattern where all users move in a straight line starting from [0 m, 750 m] coordinate in the topology and ending at [2000 m, 750 m] coordinate in the topology. Each user is located 1 meter apart from the user in front and behind that user (except for the first and last user). (ii) Random waypoint movement pattern where all users move throughout the topology by picking a destination based on generating uniformly distributed random waypoints. Each user moves at a constant speed of 2 mph. (iii) Random waypoint movement pattern where each user selects a speed in the range [2,20] mph according to a uniform distribution while moving between the current waypoint and the next waypoint.

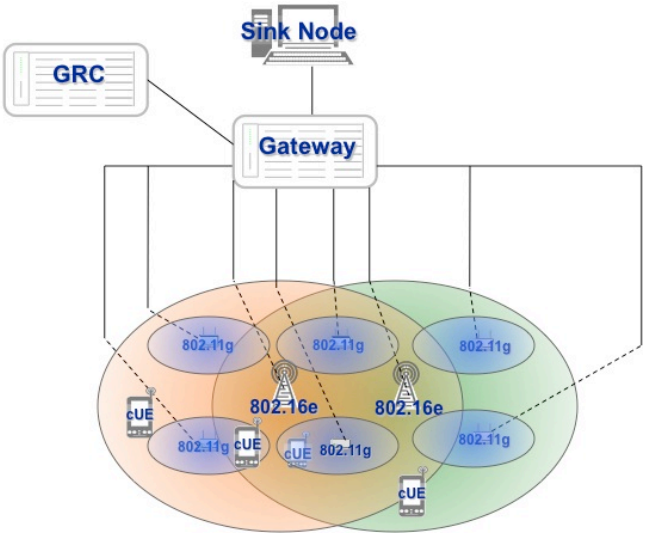

Figure 5.3: Simulation topology in ns-2

We study two variants of user device-to-AP/BS association decision solutions: distributed and centralized. In the distributed association decision solution, each device makes its own RAT association decision. For the distributed approach, each user picks a RAT according to the norm today: use a Wi-Fi network if available; otherwise, use WiMAX network. When the user is connected to WiMAX network, the user performs a Wi-Fi link scan every 5 seconds using its unused radio. If a Wi-Fi network is detected, the user starts using Wi-Fi network. When the user is connected to Wi-Fi network, the user is satisfied and does not perform any link scans. If the user receives a link going down (or link down) MIH event while using either Wi-Fi or WiMAX network, the corresponding radio goes into scan mode to search for other available Wi-Fi APs or WiMAX BSs.

For the centralized approach, each user connects to the AP/BS according to the decision made by the GRC. Each user periodically (on a 5-second basis) sends link parameter report to the GRC to inform the GRC of the available APs/BSs and the

associated MCS the user can use to connect to those APs/BSs. To obtain the link parameter status for each available AP/BS, the user device employs one of the following two solutions: periodic scanning (on a 5-second basis) on both its radios to search for Wi-Fi APs and WiMAX BSs, or location-based solution where the user sends its current location in the link parameter report. The GRC maintains a database of information related to MCS achievable with all available AP/BS by the user device at any given location. Using the link parameter report, the GRC computes the re-association decisions on a 5-second basis based on the sort-based heuristic algorithms presented in the next section. Note that if a link parameter report packet for any user device is lost (which can happen due to collisions if the packet is transmitted via a Wi-Fi connection), the GRC uses the most recent link parameter report it obtained successfully from that user device. Also, for the centralized association decision solution, if any user device receives a link going down (or link down) trigger, it does not wait for the next GRC re-association computation (and the subsequent report) to switch APs/BS. It automatically goes into scan mode on both radio interfaces and establishes a connection with an available AP/BS.

5.3.1 Greedy Sort-based Resource Allocation Algorithms

The GRC uses the link parameter report and independent scheduling objective of both WiMAX and Wi-Fi RATs when computing user device-to-AP/BS association decisions each scheduling interval. From the link parameter report, the GRC can identify the maximum data rate each user can achieve using any BS/AP (depending on the MCS/location the user reported). The mapping of each MCS to the maximum data rate for both Wi-Fi and WiMAX RATs is available in Appendix C. The information related to

maximum data rate that each user can achieve via all available AP/BS is used by GRC in computing the user device-to-AP/BS associations. The pseudo-code for the heuristic algorithms used by the GRC is presented in Appendix E.

The first heuristic algorithm tries to maximize system spectral efficiency and the second heuristic algorithm tries to maximize global instantaneous fairness. Each algorithm first sorts each user in descending order for each AP/BS based on the maximum data rate the user can achieve via the corresponding AP/BS. In case of ties, the user with lowest achievable overall data rate over all APs/BSs (and hence having fewer options) is put ahead of the other tied users. Based on this sorted order, in each decision round both algorithms compute the achievable total system throughput and lowest user throughput metrics under the assumption that the best unassociated user for each AP/BS is associated to the corresponding AP/BS. In performing these computations, the GRC uses the scheduling objective of each AP/BS (proportional fairness or max-min fairness) to determine the percentage of air-time usage (p_{ua}) user $u \in U$ gets through AP/BS $a \in A$ if the next best unassociated user $u \in U$ is associated to AP/BS $a \in A$. p_{ua} is determined according to Proposition 5.1 or Proposition 5.2 if the scheduling objective of AP/BS $a \in A$ is proportional fairness or max-min fairness respectively. Using p_{ua}, the values for total throughput through AP/BS $a \in A$ and lowest user throughput are computed using equations presented in lines 23 and 24 of the pseudo-code respectively if the scheduling objective of AP/BS $a \in A$ is max-min fairness and the values for total throughput through AP/BS $a \in A$ and lowest user throughput are computed using equations presented in lines

27 and 28 of the pseudo-code respectively if the scheduling objective of AP/BS $a \in A$ is proportional fairness.

Proposition 5.1: For a single independent multi-rate network $a \in A$, proportional fairness is achieved when the percentage of air-time usage (p_{ua}) of all users $u \in U$ connected to network a (represented by $u \in Ua$) is equal, i.e. $p_{ua} = \frac{1}{|Ua|}$.

Proof: Presented in Appendix F.

Proposition 5.2: For a single independent multi-rate network $a \in A$, max-min fairness is achieved when the percentage of air-time usage (p_{va}) of user $v \in U$ connected to network a (represented by $v \in Ua$) is given by $p_{va} = \frac{1}{\sum_{u \in Ua}\left(\frac{r_{va,max}}{r_{ua,max}}\right)}$.

Proof: Presented in Appendix F.

Based on the achievable total throughput through AP/BS $a \in A$ and lowest user throughput computations made for each AP/BS $a \in A$ under the assumption that the next best unassociated user $u \in Ua$ is connected to AP/BS $a \in A$, each heuristic algorithm makes its next user device-to-AP/BS association decision according to lines 31-49 presented in the pseudo-code. The first heuristic algorithm trying to maximize system spectral efficiency makes decisions based on maximum achievable total system throughput and the second heuristic algorithm trying to maximize instantaneous fairness makes decisions based on maximum achievable lowest user throughput. In case of ties, each algorithm makes decision based on the other metric (maximum lowest user throughput metric for algorithm trying to maximize system spectral efficiency and maximum achievable total system throughput metric for algorithm trying to maximize

instantaneous fairness) to break ties. The process of making user device-to-AP/BS association decisions based on the computed achievable total system throughput and lowest user throughput metrics in each decision round continues until all users are associated to a AP/BS. We use the term *Max Throughput* algorithm for the first algorithm trying to maximize system spectral efficiency and *Max Fairness* algorithm for the second algorithm trying to maximize instantaneous fairness in the remainder of our work.

Example: We provide an illustrative example that further clarifies the association decisions made by the two centralized algorithms in each round. Consider a hetnet system with two BSs (a and b) shown in Figure 5.4. BS a implements an independent max-min fairness scheduler and BS b implements an independent proportional fairness scheduler. There are four users in the hetnet system and each user can achieve a maximum data rate via BS a and BS b shown in Figure 5.4. The first step for both centralized algorithms sorts each user for both BSs based on the maximum achievable data rates as shown in Table 5.4. The second step uses the sorted order presented in Table 5.4 to compute a user device-to-BS association decision based on maximum achievable total system throughput (for Max Throughput algorithm) and maximum lowest user throughput (for Max Fairness algorithm) metrics during each round. The association decision for each round for Max Throughput and Max Fairness algorithms is presented in Tables 5.5 and 5.6 respectively. The metric used to make the decision during each round is colored red in the corresponding table. In case of ties, the first metric is colored green and the second metric used to make the decision is colored red.

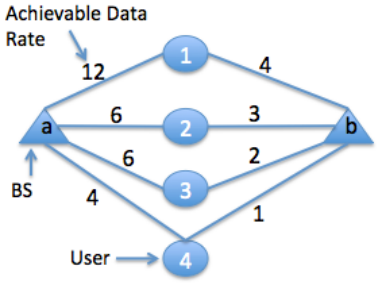

Figure 5.4: Example hetnet scenario

Table 5.4: Sorted user for each BS

BS a Max-Min Fair	BS b Proportional Fair
User 1	User 1
User 3	User 2
User 2	User 3
User 4	User 4

Table 5.5: Max throughput algorithm association decisions

Round #	Total System Throughput (Next assoc. BS a)	Lowest User Throughput through BS a	Total System Throughput (Next assoc. BS b)	Lowest User Throughput through BS b	Association Decision
1	12	12	4	4	User 1 – BS a
2	8	4	15	3	User 2 – BS b
3	11	4	14.5	1	User 3 – BS b
4	8.5	3	14	0.333	User 4 – BS b

Table 5.6: Max fairness algorithm association decisions

Round #	Total System Throughput (Next assoc. BS a)	Lowest User Throughput through BS a	Total System Throughput (Next assoc. BS b)	Lowest User Throughput through BS b	Association Decision
1	12	12	4	4	User 1 – BS a
2	8	4	15	3	User 3 – BS a
3	7.2	2.4	11	3	User 2 – BS b
4	9	2	10	2	User 4 – BS b

5.3.2 Greedy Sort-based Algorithm Results and Analysis

We first assess the benefits of the centralized solution implemented at the GRC compared to the distributed solution in terms of achieved system spectral efficiency. Each

simulation is run in ns-2 for 2000 seconds. The pattern for achieved results in terms of all network efficiency measures for both WiMAX MAC implementations (proportional fairness and max-min fairness objectives) is similar. Hence, we provide results using the proportional fairness WiMAX MAC implementation for the remainder of our work. Note that proportional fairness WiMAX MAC achieves higher throughput (and spectral efficiency) at the expense of instantaneous and long-term fairness. Also note that Wi-Fi MAC implementation is set to the default IEEE 802.11g behavior and is always assumed to achieve max-main fairness in our solution. The spectral efficiency comparisons for each solution combination (centralized or distributed decision making and each resource allocation procedure) are presented in Figure 5.5. The results shown in Figure 5.5 present the average system spectral efficiency averaged over entire simulation duration. Recall that spectral efficiency metric for each scheduling interval t is computed using (3.1) presented in Chapter 3.

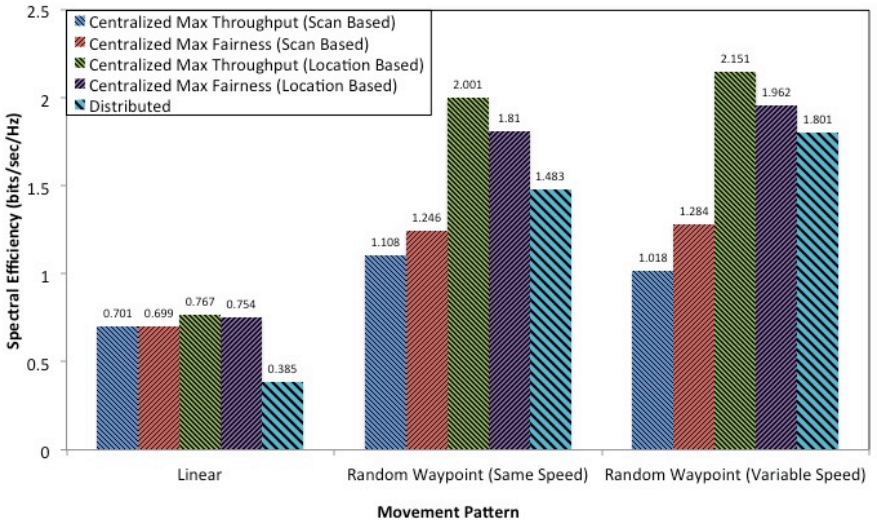

Figure 5.5: Spectral efficiency comparisons

As can be seen from Figure 5.5, the location based centralized solutions (with both Max Throughput and Max Fairness resource allocation procedures) outperform the distributed solution in terms of spectral efficiency for all movement patterns due to the benefits of multi-access network diversity. The gain in spectral efficiency for the centralized Max Throughput resource allocation procedure compared to the distributed solution is 99.2% (from 0.385 bits/sec/Hz to 0.767 bits/sec/Hz), 34.9% (from 1.483 bits/sec/Hz to 2.001 bits/sec/Hz) and 19.4% (from 1.801 bits/sec/Hz to 2.151 bits/sec/Hz) for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. The gain in spectral efficiency for the centralized Max Fairness resource allocation procedure compared to the distributed solution is 95.8% (from 0.385 bits/sec/Hz to 0.754 bits/sec/Hz), 22.0% (from 1.483 bits/sec/Hz to 1.81 bits/sec/Hz) and 8.9% (from 1.801 bits/sec/Hz to 1.962 bits/sec/Hz) for the linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. As expected, the gain in spectral efficiency is higher for the resource allocation algorithm trying to maximize system spectral efficiency (as compared to instantaneous fairness). The highest gain in spectral efficiency for a centralized solution (for both resource allocation procedures) occurs for the linear movement pattern where all the users are grouped together and experience similar connectivity conditions and the lowest gain occurs for the random waypoint (variable speed) movement pattern where all the users (because of the randomness in their movement patterns) experience the most frequent change in connectivity conditions. For linear movement pattern, all users experience similar connectivity conditions (for example, one Wi-Fi AP and one WiMAX BS is available to

all users at the same time), and as a result the distributed algorithm performs very poorly as each user for this solution will select the Wi-Fi access network for data connectivity. The centralized solution (both resource allocation algorithms) intelligently associates some users to Wi-Fi AP and other users to WiMAX BS and as a result achieves significant performance improvement. Also, for the linear movement pattern, since all users are grouped together, resources of only one Wi-Fi AP are used at any given time in addition to the two WiMAX BSs. Whereas for the random waypoint movement pattern, since all users are spread out throughput the network topology, up to six Wi-Fi APs are used at any given time in addition to the two WiMAX BSs. As a result, the overall spectral efficiency obtained for the linear movement pattern (for any association decision solution) is much lower than that of random waypoint movement pattern, which can be seen in Figure 5.5.

The scan based centralized solution has technical challenges associated with it. For this solution, since the radios on the user device disrupt active data connections to search for available networks on a periodic basis (5 seconds), multiple data packets sent by the sink node are either dropped or significantly delayed. This phenomenon results in a TCP timeout and resetting (halving) of the window size by the AIMD TCP congestion control mechanism every 5 seconds. As a result, there usually aren't enough data packets at the BS/AP to send to each connected user to fully utilize the radio link. Moreover, if all users connected to the BS/AP scan at the same time, no traffic is sent by the corresponding BS/AP for the scan duration resulting in further underutilization of the radio link. So, the performance achieved by this solution is quite unpredictable as it

depends heavily on the scanning process. While the centralized scan solution (both resource allocation algorithms) for linear movement pattern outperforms the distributed solution as seen from Figure 5.5, this solution performs worse than the distributed solution for both random waypoint movement patterns. To remedy this challenge, a solution needs to be worked on where active data connections are not disrupted during the scanning process. However, this requires extra dedicated hardware on user devices for scanning purposes or the information related to neighbor APs/BSs needs to be broadcasted to the user devices on a periodic basis by the serving AP/BS in a smart fashion. We omit results related to the scan based centralized solution for the remainder of our work.

We next present the instantaneous and long-term fairness results in Figures 5.6 and 5.7 respectively. Since we only consider best-effort traffic, the instantaneous fairness metric for each scheduling interval is computed using (3.5) provided in Chapter 3 and the results averaged over the entire simulation run are presented in Figure 5.6. The long-term fairness results presented in Figure 5.7 are computed using (3.3) provided in Chapter 3. The results for both instantaneous and long-term fairness follow the same trend. As can be seen from Figures 5.6 and 5.7, the location based centralized solution, which attempts to maximize instantaneous fairness, outperforms the distributed solution in terms of both instantaneous and long-term fairness. The gain in instantaneous fairness metric for the centralized Max Fairness resource allocation procedure compared to the distributed solution is 12.9% (from 0.769 to 0.868), 28.5% (from 0.312 to 0.401) and 8.0% (from 0.275 to 0.297) for linear, random waypoint same speed and random waypoint variable

speed movement patterns respectively. The gain in long-term fairness metric for the centralized Max Fairness resource allocation procedure compared to the distributed solution is 1.8% (from 0.981 to 0.999), 9.3% (from 0.691 to 0.755) and 2.4% (from 0.777 to 0.796) for the linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. Note that this improvement for the centralized Max Fairness resource allocation procedure compared to the distributed solution for both fairness metrics is experienced in addition to the spectral efficiency improvement shown for this procedure in Figure 5.5. So the centralized solution with Max Fairness resource allocation procedure improves both conflicting objectives of maximizing system throughput and (instantaneous and long-term) fairness compared to a distributed solution by making smart association decisions reaping the benefits of multi-access network diversity. Both instantaneous and long-term fairness metrics for the centralized Max Throughput resource allocation procedure suffer compared to the distributed solution for all user movement patterns as seen from Figures 5.6 and 5.7 respectively. But the degradation in fairness metrics for the centralized solution with Max Throughput resource allocation procedure comes as a cost of achieving highest system spectral efficiency as shown in Figure 5.5.

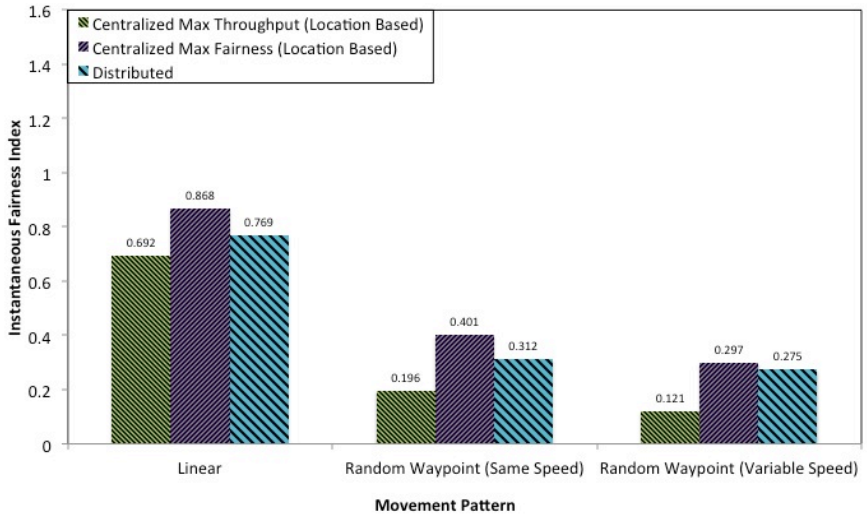

Figure 5.6: Instantaneous fairness comparisons

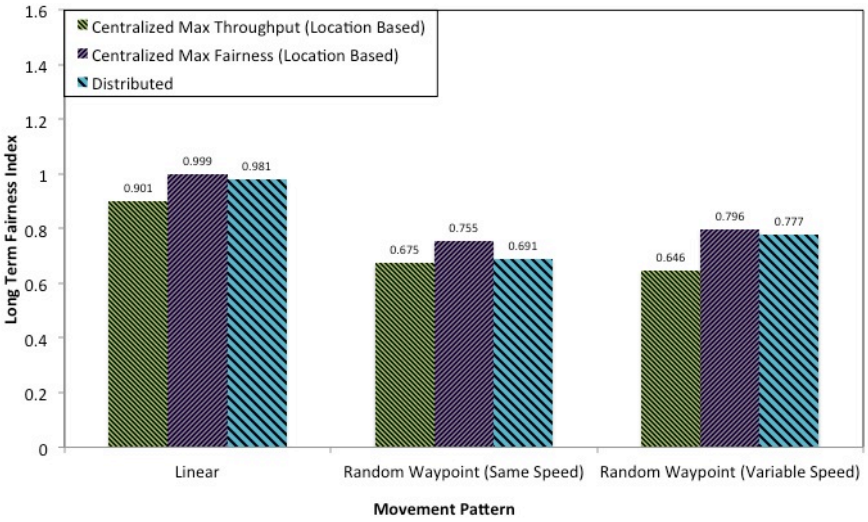

Figure 5.7: Long-term fairness comparisons

We finally present the power consumption results in Figure 5.8. The energy consumption computation follows the same approach presented in (3.7) in Chapter 3, where the energy consumption of a user device depends on two components: the number of bits transmitted/received using Wi-Fi/WiMAX RAT and the number of handovers performed by the device. For the first component, we use the same $E_{t,a}$ numbers for Wi-Fi and WiMAX RATs as presented in Table 3.3. But the second component ($E_{o,a}$), which

represents the overhead in terms of energy consumption during a handover, changes for this study because of our assumptions. In (3.7), we only model reconfiguration handoffs and assume that $E_{o,a}$ energy is consumed when a reconfiguration handoff (RAT change) takes place and that this event requires hardware reconfiguration ($E_{rec,a}$ energy cost) as well as it has RAT association energy costs ($E_{assoc,a}$). But now, we model a horizontal handover (WiMAX-to-WiMAX) in addition to a vertical/reconfiguration handover. So, for a horizontal handover ($\mathrm{HH_{ua}}$), only the $E_{assoc,a}$ energy costs are incurred while for a vertical handover ($\mathrm{VH_{ua}}$), both $E_{rec,a}$ and $E_{assoc,a}$ energy costs are incurred. The equation to compute overall energy consumption for a user device (ω_u) during the entire simulation run is presented in (5.1), where η_{ua} represents the number of data bits transmitted by user $u \in U$ over RAT $a \in A$, $|\mathrm{HH_{ua}}|$ represents the number of horizontal handovers experienced by user $u \in U$ within RAT $a \in A$ and $|\mathrm{VH_{ua}}|$ represents the number of vertical handovers experienced by user $u \in U$ to RAT $a \in A$ during the entire simulation run. Moreover, for this study since we assume the use of static multi-modal radios (based on ASIC hardware), the $E_{rec,a}$ numbers are much lower than the ones used in our previous study in Chapter 4 where reconfigurable radio hardware is assumed. We summarize the $E_{t,a}$, $E_{rec,a}$, and $E_{assoc,a}$ energy consumption numbers used in this study in Table 5.7. The results presented in Figure 5.8 represent the average power consumption cost per user. The energy consumption, ω_u, for each user $u \in U$ is computed using (5.1). The computed ω_u value for each user is summed and the sum is divided by the simulation duration to obtain average power consumption per user.

$$\omega_u = \sum_{a \epsilon A} [\, E_{t,a}(\eta_{ua}) + |\mathrm{HH_{ua}}| * E_{assoc,a} + |\mathrm{VH_{ua}}| * (E_{rec,a} + E_{assoc,a}) \,] \qquad (5.1)$$

Table 5.7: Energy consumption components for simulated RATs

	802.11g	802.16e
$E_{t,a}$ (Joules/x KB)	0.007(x)	0.018(x)
$E_{rec,a}$ (Joules)	0.05	0.28
$E_{assoc,a}$ (Joules)	5.9	3.2

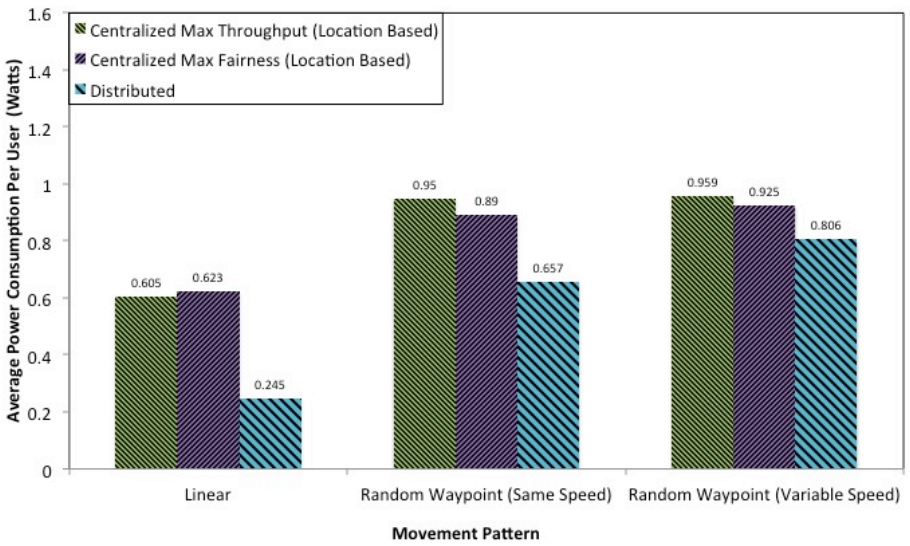

Figure 5.8: Power consumption comparisons

Since the average power consumption model depends on two components (energy consumption per bit transmitted/received and the number of handovers), the average power consumption results include the effects of both actions. As seen from Figure 5.8, the average power consumption trend generally mimics the spectral efficiency trend shown in Figure 5.5. This indicates that the first power component (energy consumption per bit transmitted/received) dominates the overall power consumption. Since the

centralized solutions achieve higher spectral efficiency (and as result transmit/receive more data bits), the overall power consumption for centralized solutions is higher than that of the distributed solution. To quantify the power consumption increase resulting from frequent re-associations for a centralized solution more accurately, the overhead results are presented next.

The number of horizontal (WiMAX-to-WiMAX) and vertical (WiMAX-to-Wi-Fi and Wi-Fi-to-WiMAX) handovers determines the energy consumed by the centralized and distributed solutions resulting from re-associations. The actual number of each type of handover for each simulation scenario is presented in Figure 5.9. As can be seen from the figure, the horizontal handovers dominate the total number of handovers for the centralized solutions in each movement pattern. There are approximately 50, 60 and 30 times more horizontal handovers for the centralized solution with Max Throughput resource allocation procedure compared to the distributed solution for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. There is a 50, 10 and 10 times increase in horizontal handovers for the centralized solution with Max Fairness resource allocation procedure compared to the distributed solution for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. As a result, the increase in energy consumption resulting from re-associations for the centralized solutions is an order of magnitude higher compared to the distributed solution. However, horizontal handovers do not consume as much energy as a vertical handover (3.2 Joules for a WiMAX-to-WiMAX handover compared to 3.48 Joules for a WiFi-to-WiMAX handover and 5.95

Joules for a WiMAX-to-Wi-Fi handover). Also, the number of horizontal handovers for a distributed solution are low compared to vertical handovers as seen in Figure 5.9. So, the impact of horizontal handovers on overall increase in energy (and subsequently power) consumption is not as significant as that of vertical handovers. The increase in vertical handovers for the centralized solution (both Max Throughput and Max Fairness resource allocation procedure) is all under a factor of 4 times greater compared to the distributed solution. Hence, vertical handovers do not cause a significant increase in power consumption. The average power consumption results that consider both horizontal and vertical handovers for each solution are presented in Figure 5.10.

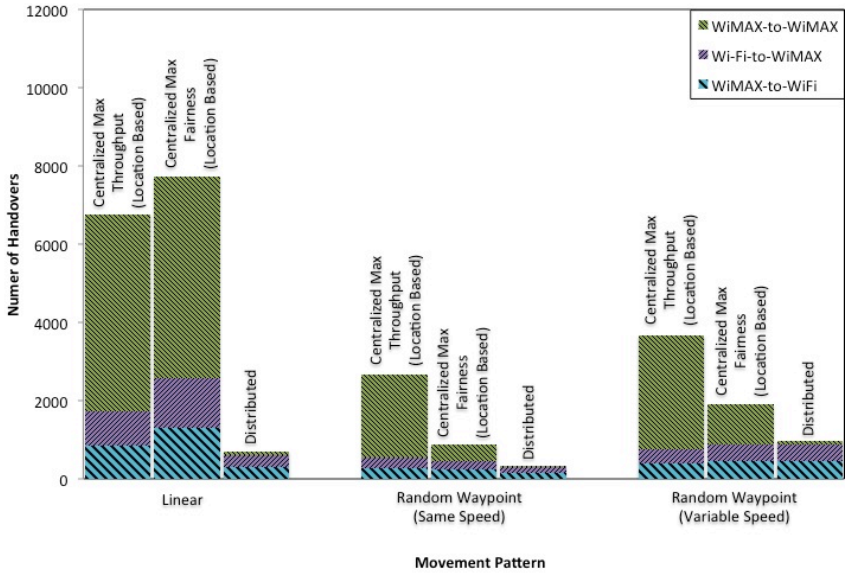

Figure 5.9: Handover comparisons

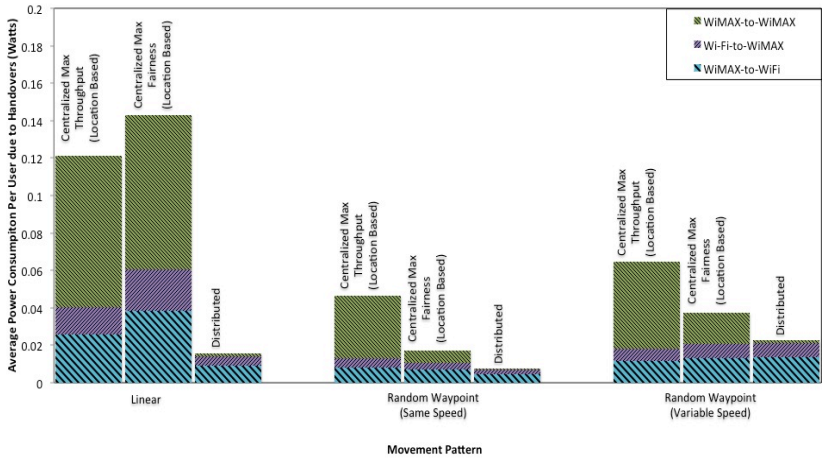

Figure 5.10: Average power consumption comparisons due to handovers

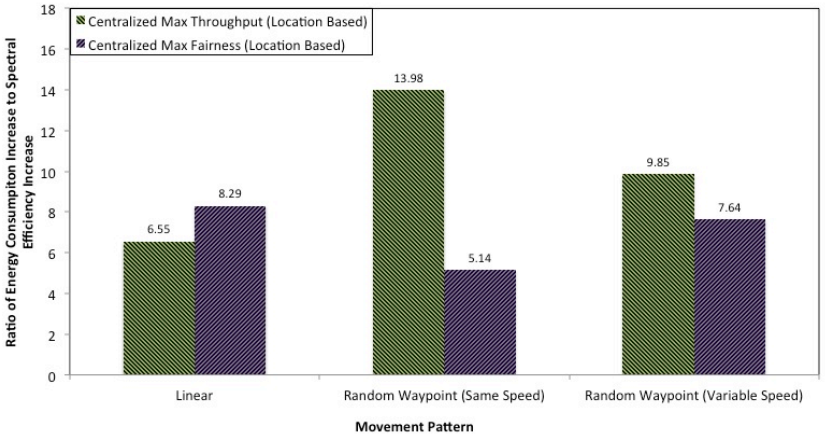

Figure 5.11: Ratio of relative increase in power consumption due to handovers compared to spectral efficiency

As seen from Figure 5.10, the increase in average power consumption per user resulting from frequent re-associations (handovers) for the centralized solution with Max Throughput resource allocation procedure is 650% (0.12 Watts compared to 0.016 Watts), 488% (0.047 Watts compared to 0.008 Watts) and 191% (0.064 Watts compared to 0.022 Watts) compared to the distributed solution for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. The increase in the same metric for the centralized solution with Max Fairness resource allocation

procedure compared to the distributed solution is 794% (0.143 Watts compared to 0.016 Watts), 113% (0.017 Watts compared to 0.008 Watts) and 68% (0.037 Watts compared to 0.022 Watts) for linear, random waypoint same speed and random waypoint variable speed movement patterns respectively. To get an estimate of the increase in power consumption (due to handovers) relative to the increase in spectral efficiency (which is in the range of [8.9%-99.2%]) shown in Figure 5.5, the ratio of increase in power consumption (due to handovers) to spectral efficiency increase for a centralized solution (with both resource allocation procedures) compared to the distributed solution is shown in Figure 5.11. As can be seen from this figure, the increase in power consumption is a factor of 5.14 to 13.98 times greater than the increase in spectral efficiency, which indicates an order of magnitude higher increase in power consumption compared to the increase in spectral efficiency. Note that the resource allocation algorithms used by the centralized solution did not consider any energy/power consumption or handover minimization metrics in generating the user device-to-AP/BS association decisions. To lower some of the adverse effects of frequent handovers such as the increase in power consumption just shown, extensions to the centralized heuristic algorithms could be made so that handovers for user devices only occur if certain performance improvement thresholds are crossed.

In addition to the system efficiency performance measures just presented, we analyze the messaging overhead required during network re-associations and compare it with achieved system throughput. We consider the technology-independent messages presented in Table 5.3 in our overhead modeling. These messages include periodic link

parameter report, association mapping message sent by the GRC based on re-association computations, router solicitation message sent by the user device to obtain an IP address once link-layer connectivity has been established, router advertisement message sent by the AP/BS in response to the router solicitation message, and the flow redirect request sent by the user device to inform the other end-point of the switch in interfaces. The comparison of average throughput consumed by the overhead messages vs. the actual average system (data) throughput is presented in Figure 5.12. The trend in the amount of overhead created by each solution mimics the number of handovers experienced by each solution (shown in Figure 5.9), which follows expectations. The highest amount of overhead throughput produced by any solution compared to the overall throughput is 18.3% (24.13 Mbps of data throughput and 5.41 Mbps of overhead throughput) for the Max Fairness centralized solution for the linear movement pattern. While this is a significant amount of overhead, this happens only in extreme cases where all users are grouped together in one location where each user can use a limited set of RATs. For this movement pattern, even the distributed solution has an overhead throughput of 15.3% (12.32 Mbps of data throughput and 2.22 Mbps of overhead throughput). For the users that are spread throughout the topology (random waypoint mobility pattern), the highest overhead throughput is 4.7% (64.03 Mbps of data throughput and 3.15 Mbps of overhead throughput) for the Max Throughput centralized solution. For all centralized solutions for the random waypoint movement patterns (same speed and variable speed), the overhead throughput to total throughput ratio is in the range [4.4%, 4.7%]. For the distributed solution for the random waypoint movement patterns (same speed and variable speed),

the overhead throughput to total throughput ratio is in the range [0.3%, 0.6%]. So as seen from these results, the overhead related to the centralized solution is very manageable and the increase in overhead due to network re-associations for a centralized solution compared to a distributed solution does not exceed by about 4.1%.

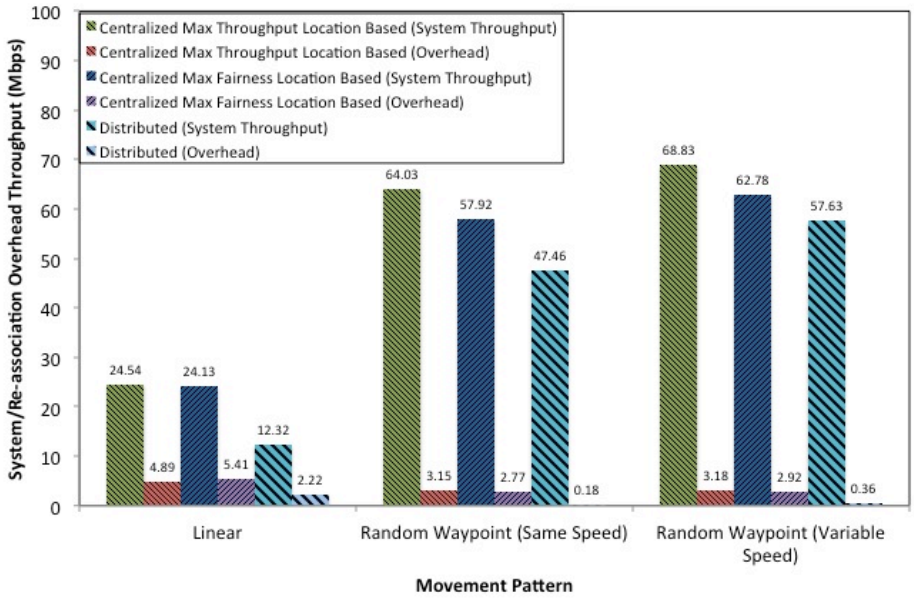

Figure 5.12: Comparison of re-association overhead vs. system throughput

To summarize the main findings of our second phase study, we identify the following conclusions:

- A hetnet based on a centralized solution can almost double its spectral efficiency (99% increase) compared to the distributed solution. There are even cases where a performance increase is achieved in both conflicting objectives of spectral efficiency and (instantaneous and long-term) fairness due to the benefits of multi-access network diversity. For our centralized solution that uses Max Fairness resource allocation procedure, using the linear user movement pattern, we showed a spectral

efficiency increase of 95.8%, instantaneous fairness increase of 12.9% and long-term fairness increase of 1.8% compared to the distributed solution.

- A centralized solution where a user device scans periodically disrupting active data connections (TCP) results in unpredictable performance results because of the TCP congestion control mechanism and the underutilization of available RATs during the scanning procedure. A location based solution such as the one we used in our study or other mechanism such as additional scanning hardware needs to be implemented at each user device to support the generation of periodic link parameter report required by a centralized hetnet solution without disrupting active data connections.
- The centralized solution experiences a significant number of handovers compared to the distributed case, and as a result there is a significant increase in power consumption (up to 794%) resulting from network re-associations for the centralized solution compared to the distributed solution. The resource allocation procedure implemented at the centralized solution needs to limit the number of handovers by using incremental policies in addition to the traditional objectives of optimizing network efficiency measures of spectral efficiency and fairness.
- The overhead required by the centralized solution based on IEEE 802.21 framework does not exceed more than 4.1% compared to a distributed solution for the various user movement patterns analyzed in our work and the overhead throughput compared to overall throughput does not exceed 18.3% (which only happens in rare cases where all users are grouped together in the linear movement pattern. The overhead

throughput accounts for fewer than 4.7% of overall throughput for all random movement patterns).

CHAPTER SIX

CONCLUSIONS

We explored radio resource allocation and management issues related to a large-scale hetnet wireless system made up of several RATs that collectively provide a unified wireless network to a diverse set of users through co-ordination managed by a centralized GRC. We have assumed centralized means 'locally centralized' where decisions are based on various amounts of information related to the mobile users and RAT systems in specific geographic vicinity. We characterized the network performance in terms of various conflicting network efficiency objectives that incorporated costs associated with a network re-association operation. We accounted for RAT-specific implementation details and the management overhead associated with setting up a centralized control.

For the first phase of our research study, using MATLAB-based simulation that uses a heuristic resource allocation algorithm that tries to maximize spectral efficiency while maintaining acceptable levels of fairness, we showed possible gains in spectral efficiency due to multi-access network diversity at the cost of increase in power consumption for two network topologies: *Balanced Topology*, where the number of RATs accessible to users of two different cellular carriers (or operators) is similar and *Unbalanced Topology*, where the number of APs/BSs accessible to users of one cellular operator is far greater compared to the users of the other operator. Our results suggest that the gains are not as significant for the balanced network topology, where the spectral efficiency increases in the range [14.4%, 75.0%]. Depending on user device (hardware)

assumptions, the corresponding increase in power consumption is in the range [114.0%, 916.8%] or [32.2%, 129.8%] if the devices are completely manufactured with FPGA fabric or with ASIC components respectively. The reconfiguration rate (or number of re-associations/reconfigurations required by each user) mainly dictates the power consumption trends for balanced network deployment. For the unbalanced network topology, significant gain in spectral efficiency in the range [314.3%, 553.7%] is achieved. The corresponding increase in power consumption is in the range [104.9%, 614.9%] or [35.1%, 98.8%] if the devices are completely manufactured with FPGA fabric or with ASIC components respectively. The actual power consumption for unbalanced network deployment not only depends on the reconfiguration rate, but also on the number of RATs that are used. In the worst case (for balanced network topology), using completely reconfigurable devices (manufactured with FPGA fabric) results in almost an order of magnitude tradeoff between spectral efficiency (which increases in the range [14.4%, 75.0%]) and power consumption efficiency (which increases in the range [32.2%, 916.8%]) metrics. Moreover, the number of reconfigurable radios required per user device to achieve the increase in spectral efficiency is surprisingly low. For the simulation scenarios analyzed in the first phase of our research study, any user device utilizes two or fewer radios at any given time.

To characterize and achieve a desired tradeoff in terms of all the network efficiency measures, we then performed an optimization-based study using balanced network topology and reconfigurable user device assumption where we considered a multi-attribute optimization function consisting of spectral efficiency, battery lifetime of

each user (or overall energy consumption), and instantaneous and long-term fairness attributes for each user in the system. To compute the relative importance of each attribute, we used the Analytical Hierarchy Process (AHP) that took interview responses from wireless network providers as input and generated final weight assignments for each attribute in our optimization problem. Using the well-known utility function-based problem formulation, we showed an increase in a multi-attribute system utility measure of up to 56.7% for our algorithm compared to other widely studied resource allocation algorithms including max-sum rate, proportional fairness, max-min fairness and min power.

For the second phase of our research study, we used detailed ns-2 modeling to account for implementation details and overhead associated with the proposed centralized solution (GRC) in a hetnet system. We implemented two variants of sort-based user device-to-AP/BS heuristic association algorithms that considered the network performance objectives of maximizing spectral efficiency and instantaneous fairness respectively. Through ns-2 simulations, we showed an increase in spectral efficiency of up to 99% and an increase in instantaneous fairness of up to 28.5% for each respective algorithm implemented at the GRC for a centralized solution compared to a distributed solution where each user makes his/her own association decision. The efficiency increase for each respective attribute comes at the cost of an order of magnitude increase in power consumption of up to 650% and 794% for each respective algorithm implemented at the GRC compared to a distributed solution because of frequent re-associations. Also, periodic scanning required by a centralized solution that disrupt active (TCP) data

connections result in unpredictable network performance. To generate periodic link parameter report, solutions that predict the maximum achievable data rate for each user using all available APs/BSs (such as a location-based strategy) is required.

In both phases of our research, we consider a specific region that is managed by a GRC. Note that for large-scale hetnet systems, several centralized controllers (GRCs) that manage different regions of overall network topology could be created in a hierarchical fashion. For one region managed by a GRC, our results from second research phase suggest that the overhead created by a centralized system is manageable (under 4.7% overhead throughput for random user distributions in the region).

Both phases of our work suggest that a significant increase in power consumption (on the order of a factor of 2 to 7) is required to achieve an increase in spectral efficiency. In fact, as illustrated in Figure 5.11, the power consumption grows an order of magnitude higher compared to the increase in spectral efficiency. This phenomenon results due to periodic user device-to-AP/BS re-associations coordinated by the GRC. Advanced power management schemes for user devices that are more appropriate for hetnet systems can reduce this power consumption. In our work, using the optimization-based resource allocation study, we showed that resource allocation algorithms implemented at the GRC could achieve a desired trade-off between spectral efficiency and energy (or power) consumption. An area of future work will blend a global allocation strategy that takes current device battery levels into account (for example, reduce the frequency of handovers as a device's battery depletes, or add elements of deferred transmissions in hopes of higher efficiency transfers in the near future).

The definitions and associated research in wireless hetnets has evolved significantly over the last several decades. Current research (such as the recent work by Andrews [109]) provides an information theoretic perspective as it tries to find the optimal fraction of traffic to offload to maximize SINR and/or data rates. Much of the recent focus is carrier centric where the core hetnet (involving pico and femto cells) is likely to be under the control of a single carrier. Our direction has been more towards an Internet model where standard protocols allow users to view a unified wireless access network that is built on any number of independent AWSs. We recognize that economic models must be considered that provide incentives for AWSs to cooperate. This issue represents further future work.

APPENDICES

Appendix A

Maximum Achievable Data Rates for RATs

Table A-1: Simulation Parameters for IEEE 802.11g (22 MHz)

Modulation and Coding Scheme	Bits/ Resource Block	Resource Blocks/ Sec	Maximum Achievable Data Rate (Mbps)
BPSK 1/11	91	9000	0.82
QPSK 1/11	182	9000	1.64
BPSK 1/2	500	9000	4.50
BPSK 3/4	750	9000	6.75
QPSK 1/2	1000	9000	9.00
QPSK 3/4	1500	9000	13.50
16-QAM 1/2	2000	9000	18.00
16-QAM 3/4	3000	9000	27.00
64-QAM 2/3	4000	9000	36.00
64-QAM 3/4	4500	9000	40.50

Table A-2: Simulation Parameters for IEEE 802.16e (10 MHz)

Modulation and Coding Scheme	Bits/ Resource Block	Resource Blocks/ Sec	Maximum Achievable Data Rate (Mbps)
QPSK 1/2	48	102000	4.90
QPSK 3/4	72	102000	7.34
16-QAM 1/2	96	102000	9.79
16-QAM 3/4	144	102000	14.69
64-QAM 1/2	144	102000	14.69
64-QAM 2/3	192	102000	19.58
64-QAM 3/4	216	102000	22.03
64-QAM 5/6	240	102000	24.48

Table A-3: Simulation Parameters for LTE (10 MHz)

Modulation and Coding Scheme	Bits/ Resource Block	Resource Blocks/ Sec	Maximum Achievable Data Rate (Mbps)
QPSK 1/2	36	163000	5.89
QPSK 3/4	54	163000	8.80
16-QAM 1/2	72	163000	11.74
16-QAM 3/4	108	163000	17.60
64-QAM 1/2	108	163000	17.60
64-QAM 2/3	144	163000	23.47
64-QAM 3/4	162	163000	26.41
64-QAM 5/6	180	163000	29.34

Table A-4: Simulation Parameters for HSPA (5 MHz)

Modulation and Coding Scheme	Bits/ Resource Block	Resource Blocks/ Sec	Maximum Achievable Data Rate (Mbps)
QPSK 1/4	2.34375	384000	0.90
QPSK 1/2	4.6875	384000	1.80
QPSK 3/4	7.03125	384000	2.70
16-QAM 1/2	9.375	384000	3.60
16-QAM 3/4	14.0625	384000	5.40
16-QAM 4/4	18.75	384000	7.20
64-QAM 3/4	22.9167	384000	8.80
64-QAM 5/6	27.47395	384000	10.55

Table A-5: Simulation Parameters for EVDO (1.25 MHz)

Modulation and Coding Scheme (and Effective code rate)	Bits/ Resource Block	Resource Blocks/ Sec	Maximum Achievable Data Rate (Mbps)
QPSK 1/5 (1/48)	64	450	0.03
QPSK 1/5 (1/24)	128	450	0.06
QPSK 1/5 (1/12)	256	450	0.12
QPSK 1/5 (1/6)	512	450	0.23
QPSK 1/3 (8/49)	512	450	0.23
QPSK 1/3 (1/3)	1024	450	0.46
QPSK 1/3 (16/49)	1024	450	0.46
8-PSK 1/3 (16/49)	1536	450	0.69
QPSK 1/3 (2/3)	2048	450	0.92
16-QAM 1/3 (16/49)	2048	450	0.92
8-PSK 1/3 (2/3)	3072	450	1.40
16-QAM 1/3 (2/3)	4096	450	1.84

Appendix B

Pseudo-code for Heuristic GRC Algorithm

```
1.  for each time unit
2.     for each user u ∈ U
3.        for each radio a ∈ A
4.           user(u).radio(a).mcs = function(user(u).radio(a).distance_from_BS)
5.           user(u).radio(a).rate = function(user(u).radio(a).mcs)
6.        end for u ∈ U
7.     end for a ∈ A

8.     for each radio a ∈ A
9.        for each user u ∈ U
10.          user(u).radio(a).rank = Sort(user(u).radio(a).mcs) % Descending order
11.       end for u ∈ U
12.    end for a ∈ A

       %  Assign equal Wi-Fi AP resources to all users that can connect to it
13.    for each Wi-Fi AP a ∈ A
14.       user(u).radio(a).assigned_bw(time_unit) = (total_AP_RBs(a)/num_conn_users) *
                                                    user(u).radio(a).rate
15.    end for

16.    % Cellular Step 1 – Assign each user T_u^t  = 100K with its best radio(s)
17.    for each user u ∈ U
18.       for each cellular radio a ∈ A
19.          sorted_radio_rank(u)(a) = Sort(user(u).radio(a).rank) % Descending order
20.       end for a ∈ A

21.       for each cellular radio a ∈ A
22.          if (user(u).assigned_bw(time_unit) < 1Mbps &&
                        remaining_slots(sorted_radio_rank(u)(a) ≥ 0)
23.             if (remaining_slots(sorted_radio_rank(u)(a)) ≥ slots_required_to_reach_T_u^t)
24.                user(u).radio(sorted_radio_rank(u)(a)).slots =  slots_required_to_reach_T_u^t
25                 remaining_slots(sorted_radio_rank(u)(a)).slots = − slots_required_to_reach_T_u^t
26.             else
27.                user(u).radio(sorted_radio_rank(u)(a)).slots =
                          remaining_slots(sorted_radio_rank(u)(a))
28.                remaining_slots(sorted_radio_rank(u)(a)) = 0
29.             end if-else
30.          end if
31.       end for a ∈ A
32.    end for u ∈ U
```

```
      % Cellular Step 2 – Assign additional resources of each RAT to ζ =10 best
                          users in  increments of ε =100K until they reach a cap of 1M
33.   for each cellular RAT a ∈ A
34.       for each user u ∈ U
35.           sorted_tech_rank(a)(u) = Sort(node(u).radio(a).rank)
36.       end for u ∈ U
37.   end for a ∈ A

38.   for each cellular RAT a ∈ A
39.       while (remaining_slots(a) > 0)
40.           users_served = 0, unservable_users = 0
41.           for each user u ∈ U
42.               if (user(sorted_tech_rank(a)(u)).assigned_bw(time_unit) < 1M &&
                                    user(sorted_tech_rank(a)(u)).radio(a).mcs > 0)
43.                   if (remaining_slots(sorted_tech_rank(a)(u)) ≥ slots_required_for_additional_ε)
44.                       user(sorted_tech_rank(a)(u)).radio(a).slots = slots_required_for_additional_ε
45.                       remaining_slots(sorted_tech_rank(a)(u)) = − slots_required_for_additional_ε
46.                   else
47.                       user(sorted_tech_rank(a)(u)).radio(a).slots =
                                        remaining_slots(sorted_tech_rank(a)(u))
48.                       remaining_slots(sorted_tech_rank(a)(u)) = 0
49.                   end if-else
50.                   users_served++
51.                   if (users_served == ζ)
52.                       break
53.                   end if
54.               else
55.                   unservable_users++
56.               end else
57.           end for u ∈ U
58.           if (unservable_users == |U|)
59.               break
60.           end if
61.       end while
62.   end for a ∈ A

63.  end for each time unit
```

Appendix C

Maximum Achievable Data Rates for RATs in ns-2 Studies

Table C-1. Simulation Parameters for IEEE 802.11g (22 MHz)

Modulation and Coding Scheme	Data Bits/ OFDM Symbol	OFDM Symbols /Sec	Maximum Achievable Data Rate (Mbps)
BPSK 1/2	24	250000	6.0
BPSK 3/4	36	250000	9.0
QPSK 1/2	48	250000	12.0
16-QAM 1/2	96	250000	24.0
16-QAM 3/4	144	250000	36.00
64-QAM 2/3	192	250000	48.00
64-QAM 3/4	216	250000	54.00

Table C-2. Simulation Parameters for IEEE 802.16e (10 MHz)

Modulation and Coding Scheme	Bits/ OFDM Symbol	OFDM Symbols/ Sec	Maximum Achievable Data Rate (Mbps)
BPSK 1/2	88	44800	3.94
QPSK 1/2	184	23400	8.24
QPSK 3/4	280	23400	12.54
16-QAM 1/2	376	23400	16.84
16-QAM 3/4	578	23400	25.89
64-QAM 2/3	760	23400	34.05
64-QAM 3/4	856	23400	38.35

Appendix D

Relevant ns-2 Simulation Parameters

Table D.1: IEEE 802.11g Simulation Parameters

Parameter	Description
CWmin	15
CWmax	1023
Beacon Interval	100 ms
Max Acceptable Beacon Loss	10
RTS/CTS Mechanism	Off
Location of APs (x, y co-ordinates)	(650, 750); (650, 1250); (1000, 750); (1000, 1250); (1350, 750); (1350, 1250)
Number of Channels	11
Channel Bandwidth	22 MHz
Supported MCS	BPSK1/2, BPSK 3/4, QPSK 1/2, 16-QAM 1/2, 16-QAM 3/4,64-QAM 2/3, 64-QAM 3/4
MCS Feedback	ACK Piggyback
Coverage Range	150 meters
Propagation Model	TwoRayGround
Scan Duration	1.32 seconds (120 ms for each channel)
Scan Mode	Passive
Link Going Down Factor	1.1

Table D.2: IEEE 802.16e Simulation Parameters

Parameter	Description
Channel Bandwidth	10 MHz
Frame Duration	5 ms
Location of BSs (x, y co-ordinates)	(500, 1000); (1500; 1000)
Scheduler	Deficit Weighted Round Robin
Scan Duration	125 ms (25 frames)
Scan Iterations	1
DL:UL Ratio	3:2
Supported MCS	BPSK1/2, QPSK 1/2, QPSK 3/4, 16-QAM 1/2, 16-QAM 3/4,64-QAM 2/3, 64-QAM 3/4
MCS Feedback	CQI Channel
Coverage Range	1 kilometer
Propagation Model	TwoRayGround
Link Going Down Factor	1.1

Table D.3: Data Flow Simulation Parameters

Parameter	Description
Traffic Direction	Downlink (sink node to cUE)
Transport Protocol	TCP
TCP Flavor	Selective ACK (Sack)
TCP Congestion Control Mechanism	Additive Increase Multiplicative Decrease (AIMD)
Traffic Pattern	Constant Bit Rate
CBR Packet Size	500 Bytes
Packet Interval	0.160 ms
Traffic Generation Rate	25 Mbps

Appendix E

Greedy Sort-Based Resource Allocation Algorithms Pseudo-code

(i) Maximizing spectral efficiency heuristic algorithm (Max Throughput)

```
1.  for each AP/BS a ∈ A
2.      sorted_users_AP_BS[a].list = Sort(r_ua,max)     % Descending order; For ties, sort according to
                                                        % user u ∈ arg min {∑_{a∈A} r_ua,max}
3.      users_allocated_AP_BS[a].list = NULL
4.      num_users_AP_BS[a] = 0
5.      achieved_throughput_AP_BS[a] = 0
6.      total_throughput_AP_BS[a] = -1
7.      lowest_user_throughput[a] = -1
8.  end for

9.  num_allocated_users = 0
10. while num_allocated_users != |U|

11.     for each AP/BS a ∈ A
12.         total_throughput_AP_BS[a] = -1
13.         lowest_user_throughput[a] = -1

14.         curr_index = 0;
15.         while userid(sorted_users_AP_BS[a][curr_index]).association_status == true
16.             curr_index++
17.         end while

18.         if curr_index ≥ sorted_users_AP_BS[a].size()
19.             continue
20.         end if

            % Compute total throughput and lowest throughput for any user u if
            % one more user is added to AP/BS a depending on its scheduler type
21.         if a.scheduler == Max_Min_Fair
22.             p_va = Calculate according to Proposition 2 where Ua includes all users
                        in users_allocated_AP_BS[a].list and v = sorted_users_AP_BS[a][curr_index]
23.             total_throughput_AP_BS[a] = (num_users_AP_BS[a] + 1) * p_va *
                                    sorted_users_AP_BS[a][curr_index]. r_ua,max
24.             lowest_user_throughput[a] = total_throughput_AP_BS[a]/(num_users_AP_BS[a] + 1)
25.         else if a.scheduler == Proportional_Fair
26.             p_ua =  Calculate according to Proposition 1 where |Ua| = num_users_AP_BS[a] + 1
27.             total_throughput_AP_BS[a] = p_ua * (sorted_users_AP_BS[a][curr_index]. r_ua,max +
                                    ∑ users_allocated_AP_BS[a]. r_ua,max)
28.             lowest_user_throughput[a] = p_ua * sorted_users_AP_BS[a][curr_index]. r_ua,max
29.         end else

30.     end for
```

```
        % Decide the next user to be added to any AP/BS based on the total_throughput_AP_BS[a]
        % and lowest_user_throughput[a] computations above. Base decisions on maximum
        % achievable system throughput. In case of ties, use the fairness metric
        % (lowest_user_throughput[a])
31.     achievable_system_throughput = -1
32.     max_achievable_system_throughput = -1
33.     AP_BS_to_Allocate = -1
34.     for each AP/BS a ∈ A
35.         if total_throughput_AP_BS[a] != -1
36.             achievable_system_throughput = (∑_{A\{a}} achieved_throughput_AP_BS[a]) +
                                                total_throughput_AP_BS[a]
37.             if achievable_system_throughput ≥ max_achievable_system_throughput
38.                 if achievable_system_throughput == max_achievable_system_throughput
39.                     if lowest_user_throuhgput[a] > lowest_user_throughput[AP_BS_to_Allocate]
40.                         max_achievable_system_throughput = achievable_system_throughput
41.                         AP_BS_to_Allocate = a
42.                     end if
43.                 else
44.                     max_achievable_system_throughput = achievable_system_throughput
45.                     AP_BS_to_Allocate = a
46.                 end else
47.             end if
48.         end if
49.     end for

50.     curr_index = 0;
51.     while userid(sorted_users_AP_BS[AP_BS_to_Allocate][curr_index])
                                                        .association_status == true
52.         curr_index++
53.     end while
54.     userid(sorted_users_AP_BS[AP_BS_to_Allocate][curr_index]).association_status = true
55.     users_allocated_AP_BS[AP_BS_to_Allocate].append(sorted_users_AP_BS[curr_index])
56.     achieved_throughput_AP_BS[AP_BS_to_Allocate] =
                                    total_throughput_AP_BS[AP_BS_to_Allocate]
57.     num_users_AP_BS[AP_BS_to_Allocate]++
58.     num_allocated_users++

59. end while
```

(ii) Maximizing instantaneous fairness heuristic algorithm (Max Fairness)

```
1.  for each AP/BS a ∈ A
2.      sorted_users_AP_BS[a].list = Sort(r_ua,max)     % Descending order; For ties, sort according to
                                                        % user u ∈ arg min {∑_{a∈A} r_ua,max}
3.      users_allocated_AP_BS[a].list = NULL
4.      num_users_AP_BS[a] = 0
5.      achieved_throughput_AP_BS[a] = 0
6.      total_throughput_AP_BS[a] = -1
7.      lowest_user_throughput[a] = -1
8.  end for

9.  num_allocated_users = 0
10. while num_allocated_users != |U|

11.     for each AP/BS a ∈ A
12.         total_throughput_AP_BS[a] = -1
13.         lowest_user_throughput[a] = -1

14.         curr_index = 0;
15.         while userid(sorted_users_AP_BS[a][curr_index]).association_status == true
16.             curr_index++
17.         end while

18.         if curr_index ≥ sorted_users_AP_BS[a].size()
19.             continue
20.         end if

            % Compute total throughput and lowest throughput for any user u if
            % one more user is added to AP/BS a depending on its scheduler type
21.         if a.scheduler == Max_Min_Fair
22.             p_va = Calculate according to Proposition 2 where Ua includes all users
                       in users_allocated_AP_BS[a].list and v = sorted_users_AP_BS[a][curr_index]
23.             total_throughput_AP_BS[a] = (num_users_AP_BS[a] + 1) * p_va *
                                            sorted_users_AP_BS[a][curr_index]. r_ua,max
24.             lowest_user_throughput[a] = total_throughput_AP_BS[a]/(num_users_AP_BS[a] + 1)
25.         else if a.scheduler == Proportional_Fair
26.             p_ua =  Calculate according to Proposition 1 where |Ua| = num_users_AP_BS[a] + 1
27.             total_throughput_AP_BS[a] = p_ua * (sorted_users_AP_BS[a][curr_index]. r_ua,max +
                                            ∑ users_allocated_AP_BS[a]. r_ua,max)
28.             lowest_user_throughput[a] = p_ua * sorted_users_AP_BS[a][curr_index]. r_ua,max
29.         end else

30.     end for
```

```
        % Decide the next user to be added to any AP/BS based on the lowest_user_throughput[a]
        % and total_throughput_AP_BS[a] computations above. Base decisions on trying to
        % maximize lowest throughput achieved by any user. In case of ties, use the
        % total_throughput_AP_BS[a] metric.
31.      achievable_lowest_user_throughput = -1
32.      AP_BS_to_Allocate = -1
33.      for each AP/BS a ∈ A
34.          achievable_system_throughput[a] = -1
35.          if lowest_user_throughput[a] != -1
36.              achievable_system_throughput[a] = (∑_{A\{a}} achieved_throughput_AP_BS[a]) +
                                                   total_throughput_AP_BS[a]
37.              if lowest_user_throughput[a] ≥ achievable_lowest_user_throughput
38.                  if lowest_user_throughput[a] == achievable_lowest_user_throughput
39.                      if achievable_system_throghput[a]  >
                                          achievable_system_throughput[AP_BS_to_Allocate]
40.                          achievable_lowest_user_throughput = lowest_user_throughput[a]
41.                          AP_BS_to_Allocate = a
42.                      end if
43.                  else
44.                      achievable_lowest_user_throughput = lowest_user_throughput[a]
45.                      AP_BS_to_Allocate = a
46.                  end else
47.              end if
48.          end if
49.      end for

50.      curr_index = 0;
51.      while userid(sorted_users_AP_BS[AP_BS_to_Allocate][curr_index])
                                                                .association_status == true
52.          curr_index++
53.      end while
54.      userid(sorted_users_AP_BS[AP_BS_to_Allocate][curr_index]).association_status = true
55.      users_allocated_AP_BS[AP_BS_to_Allocate].append(sorted_users_AP_BS[curr_index])
56.      achieved_throughput_AP_BS[AP_BS_to_Allocate] =
                                                   total_throughput_AP_BS[AP_BS_to_Allocate]
57.      num_users_AP_BS[AP_BS_to_Allocate]++
58.      num_allocated_users++

59.  end while
```

Appendix F

Air-time Usage Proofs for Proportional Fairness and Max-Min Fairness

Proposition 5.1: For a single independent multi-rate network $a \in A$, proportional fairness is achieved when the percentage of air-time usage (p_{ua}) of all users $u \in U$ connected to network a (represented by $u \in Ua$) is equal, i.e. $p_{ua} = \frac{1}{|Ua|}$.

Proof:

The objective of proportional fairness resource allocation problem is as follows:

$$\max \sum_{u \in Ua} \ln r_{ua}$$

$$= \max \sum_{u \in Ua} \ln (r_{ua,max} * p_{ua})$$

$$= \max \ln \prod_{u \in Ua} (r_{ua,max} * p_{ua})$$

$$= \max \prod_{u \in Ua} (r_{ua,max} * p_{ua})$$

Since $r_{ua,max} > 0$ and is a constant, the above objective reduces to the following:

$$\max \prod_{u \in Ua} p_{ua}$$

The constraints related to the optimization problem are as follows:

$$\sum_{u \in Ua} p_{ua} = 1$$

$$p_{ua} \geq 0, \forall u \in Ua$$

For solving the optimization problem given by the objective function and the two constraints above, we can successfully ignore the inequality constraint since that

constraint is non-binding as setting any $p_{ua} = 0$ would give an objective function value of 0 which would clearly not provide the maximum.

So, using the objective function and the equality constraint, we use method of Lagrangian Multipliers to solve the problem.

$$L(p_{1a}, p_{2a}, \dots, p_{|Ua|a}, \lambda) = \prod_{u \in Ua} p_{ua} + \lambda(1 - \sum_{u \in Ua} p_{ua})$$

Setting the gradient $\nabla L(p_{1a}, p_{2a}, \dots, p_{|Ua|a}, \lambda) = 0$, we get

$$\frac{\partial L}{\partial p_{1a}} = \prod_{u \in Ua, u \neq 1} p_{ua} - \lambda = 0$$

$$\frac{\partial L}{\partial p_{2a}} = \prod_{u \in Ua, u \neq 2} p_{ua} - \lambda = 0$$

.
.

$$\frac{\partial L}{\partial p_{|Ua|}} = \prod_{u \in Ua, u \neq |Ua|a} p_{ua} - \lambda = 0$$

Solving these set of equations yields

$$\prod_{u \in Ua, u \neq 1} p_{ua} = \prod_{u \in Ua, u \neq 2} p_{ua} = \cdots = \prod_{u \in Ua, u \neq |Ua|a} p_{ua} = \lambda$$

which implies

$$p_{1a} = p_{2a} = \cdots = p_{|Ua|a}$$

Using this result along with the original constraint $\sum_{u \in U} p_{ua} = 1$ results in $p_{1a} = p_{2a} = \cdots = p_{|Ua|a} = \frac{1}{|Ua|}$

■

Proposition 5.2: For a single independent multi-rate network $a \in A$, max-min fairness is achieved when the percentage of air-time usage (p_{va}) of user $v \in U$ connected to network a (represented by $v \in Ua$) is given by $p_{va} = \frac{1}{\sum_{u \in Ua}\left(\frac{r_{va,max}}{r_{ua,max}}\right)}$.

Proof:

Since we consider a single independent network, there exists only one bottleneck link. For a single bottleneck link, the max-min fairness objective results in equal data rate allocation to each user. Therefore, we obtain the following objective to provide max-min fairness:

$$r_{1a} = r_{2a} = \ldots = r_{|Ua|a}$$

$$\therefore r_{1a,max} * p_{1a} = r_{2a,max} * p_{2a} = \cdots = r_{|Ua|a,max} * p_{|Ua|a}$$

Solving p_{2a},…, $p_{|Ua|a}$ in terms of p_{1a} yields:

$$p_{ua} = \frac{r_{1a,max}}{r_{ua,max}} p_{1a} \quad \forall u \in Ua \tag{F.1}$$

Again, the constraints related to the optimization problem are as follows:

$$\sum_{u \in Ua} p_{ua} = 1$$

$$p_{ua} \geq 0, \forall u \in Ua$$

Using the value of p_{ua} in terms of p_{1a} obtained from the objective function and the first constraint, we get

$$p_{1a} \left(\sum_{u \in Ua} \frac{r_{1a,max}}{r_{ua,max}} \right) = 1$$

$$p_{1a} = \frac{1}{\sum_{u \in Ua} \frac{r_{1a,max}}{r_{ua,max}}}$$

We obtain the required values of p_{va} for any user $v \in Ua$ by using the relationship between p_{ua} and p_{1a} presented in (F.1). That is,

$$p_{va} = \frac{1}{\sum_{u \in Ua} \frac{r_{va,max}}{r_{ua,max}}}$$

■

REFERENCES

[1] Gartner Research, "Gartner Highlights Key Predictions for IT Organizations and Users in 2010 and Beyond," Available Online: http://www.gartner.com/ newsroom/id /1278413, Last accessed July 2013.

[2] Pew Research Center, "Global Digital Communication: Texting, Social Networking Popular Worldwide," Pew Report, Available Online: www.pewglobal.org/files/ 2011/12/Pew-Global-Attitudes-Technology-Report-FINAL-December-20-2011.pdf, Last accessed July 2013.

[3] Aaron Smith, "Smartphone Usage Report," Available Online: www.pewinternet. org/~/ media//Files/Reports/2011/PIP_Smartphones.pdf, Last accessed July 2013.

[4] Wireless Broadband Alliance, "Global Trends in Public Wi-Fi," Available Online: www.wballiance.com/resource-centre/global-developments-wifi-report.html, Last accessed July 2013.

[5] Beecham Research, "Worldwide Cellular M2M Modules Forecast Market Brief," August 2010.

[6] S. Hachem, T. Teixeira, V. Issarny, "Ontologies for the Internet of Things," Proceedings of the 8th Middleware Doctoral Symposium, December 2011.

[7] Y. Chen, "Challenges and Opportunities of Internet of Things," Proceedings of the Design Automation Conference (ASP-DAC), January 2012.

[8] FCC, "Mobile Broadband: The Benefits of Additional Spectrum," October 2010.

[9] Mobile Future Engineering Analysis, "The Spectrum Imperative: Mobile Broadband Spectrum and its Impacts for U.S. Consumers and the Economy: An Engineering Analysis," March 16, 2011.

[10] FCC, ET Docket No 03-222 Notice of Proposed Rule Making and Order, December 2003.

[11] A. Nosratinia, T.E. Hunter, A. Hedayat, "Co-operative Communication in Wireless Networks," IEEE Communications Magazine, vol.42, no.10, Oct. 2004.

[12] G. Kramer, M.Gastpar, P.Gupta, "Co-operative Strategies and Capacity Theorems for Relay Networks," IEEE Transactions on Information Theory, vol.51, no.9, pp. 3037-3063, Sep. 2005.

[13] A. Host-Madsen, "Capacity Bounds for Co-operative Diversity," IEEE Transactions on Information Theory, vol.52, no.4, pp.1522-1544, April 2006.

[14] C. Santivanez, R. Ramanathan, C. Partridge, R. Krishnan, M. Condell, S. Polit, "Opportunistic Spectrum Access: Challenges, Architecture, Protocols," WiCon, August 2006.

[15] Y. Xing, R. Chandramouli, S. Mangold, S. Shankar, "Dynamic Spectrum Access in Open Spectrum Wireless Networks," IEEE Journal on Selected Areas in Communications, vol. 24, no. 3, pp 626-637, March 2006.

[16] V. R. Petty, R. Rajbanshi, D. Datla, W. Frederick, D. Daniel, et. al, "Feasibility of Dynamic Spectrum Access in Underutilized Television Bands," in New Frontiers in Dynamic Spectrum Access Networks, April 2007.

[17] I. Akyildiz, W. Lee, M. Vuran, S. Mohanty, "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," Elsevier Computer Networks, vol. 50, pp. 2127-2159, 2006.

[18] E. Buracchini , "The Software Radio Concept," IEEE Communications Magazine, vol.38, no.9, pp.138-143, Sep. 2000.

[19] H. Wally, W. Tuttlebee, "Software Defined Radio: Enabling Technologies," John Wiley and Sons, 2002.

[20] R. Bagheri, A. Mirzaei, M. E. Heidari, S. Chehrazi, L. Minjae, M. Mikhemar, W. K. Tang, A. A. Abidi, "Software-Defined Radio Receiver: Dream to Reality," IEEE Communications Magazine, vol. 44, no. 8, pp.111-118, Aug. 2006.

[21] J. Mitola, "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio," PhD Thesis, Royal Institute of Technology (KTH), 2000.

[22] J. Mitola, "Cognitive Radio for Flexible Mobile Multimedia Communications," Mobile Networks and Applications, vol. 6, no. 5, Sep. 2001.

[23] S.Haykin, "Cognitive Radio: Brain-empowered Wireless Communications," IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, pp 201- 220, Feb. 2005.

[24] N. Devroye, P. Mitran, V. Tarokh, "Achievable Rates in Cognitive Radio Channels," IEEE Transactions on Information Theory, vol. 52, no. 5, pp. 1813-1827, May 2006.

[25] R.W. Thomas, L. A. DaSilva, A. B. MacKenzie, "Cognitive networks," Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Nov. 2005.

[26] R. Thomas, D. Friend, L. DaSilva, A. Mackenzie, "Cognitive Networks: Adaptation and Learning to Achieve End-to-end Performance Objectives," IEEE Communications Magazine, vol. 44, no. 12, December 2006.

[27] C. Fortuna, M. Mohorcic, "Trends in the Development of Communication Networks: Cognitive Networks," Computer Networks, 2009.

[28] E. Poorter, B. Latre, I. Moerman, P. Demeester, "Symbiotic Networks: Towards a New Level of Co-operation Between Wireless Networks," Wireless Personal Communications, vol. 45, no. 4, pp. 479-495, June 2008.

[29] T. Bu, L. Li, R. Ramjee, "Generalized Proportional Fair Scheduling in Third Generation Wireless Data Networks," Proceedings of IEEE INFOCOM, 2006.

[30] T-Mobile, Available Online: http://support.t-mobile.com/docs/DOC-1680, Last accessed September 2011.

[31] V. Chandrasekhar, J. Andrews, A. Gatherer, "Femotcell Networks: A Survey," IEEE Communications Magazine, vol. 46, no. 9, pp. 59-67, September 2008.

[32] Y. Shi, A. Mackenzie, L. DaSilva, K. Ghaboosi, M. Latvaho, "On Resource Reuse for Cellular Networks with Femto and Macrocell Coexistence," Proceedings of IEEE Globecom, December 2010.

[33] S. Lee, K. Sriram. K. Kim, Y. Kim, and N. Golmie, "Vertical Handoff Decision Algorithms for Providing Optimized Performance in Heterogeneous Wireless Networks," Proceedings of IEEE Transactions on Vehicular Technology, vol. 58, no. 2, Feb. 2009.

[34] P. Kosmides, A. Rouskas, and M. Anagnostou, "Network Selection in Heterogeneous Wireless Environments," Proceedings of the International Conference on Telecommunications, 2011.

[35] A. Calvagna, and G. Di Modica, "A User-Centric Analysis of Vertical Handovers," Proceedings of the Second ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots, 2004.

[36] E. Stevens-Navarro, and V. Wong, "Comparison Between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks," Proceedings of IEEE Vehicular Technology Conference, 2006.

[37] P. Chan, R. Sheriff, Y. Hu, P. Conforto and C. Tocci, "Mobility Management Incorporating Fuzzy Logic for a Heterogeneous IP Environment," IEEE Communications Magazine, vol. 39, no. 12, 2001.

[38] K. Radhika, and A. V. Reddy, "Network Selection in Heterogeneous Wireless Networks Based on Fuzzy Multiple Criteria Decision Making," International Journal of Computer Applications, vol. 22, no. 1, May 2011.

[39] W. Luo, and E. Bodanese, "Optimising Radio Access in a Heterogeneous Wireless Network Environment," Proceedings of the International Conference on Communications, 2009.

[40] B. Bakmaz, Z. Bojkovic, and M. Bakmaz, "Network Selection Algorithm for Heterogeneous Wireless Environment," Proceedings of PIMRC, 2007.

[41] G. Fodor, A. Eriksson, and A. Tuoriniemi, "Providing Quality of Service in Always Best Connected Networks," IEEE Communications Magazine, vol. 22, no. 41, 2003.

[42] M. Buddhikot, G. Chandranmenon, G. Ch, S. Han, Y. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and Third-Generation Wireless Data Networks," Proceedings of IEEE INFOCOM, 2003.

[43] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An Overview of IEEE 802.21: Media-Independent Handover Services," IEEE Wireless Communications, vol. 15, no. 4, 2008.

[44] G. Lampropoulos, A. Salkintzis, and N. Passas, "Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks," IEEE Communications Magazine, vol. 46, no. 1, 2008.

[45] Q. Mussabbir, W. Yao, "Optimized FMIPv6 Handover Using IEEE 802.21 MIH Services," Proceedings of first ACM/IEEE international workshop on Mobility in the Evolving Internet Architecture, 2006.

[46] A. Dutta, D. Famolari, S. Das, Y. Ohba, V. Fajardo, K. Taniuchi, R. Lopez, and H. Schulzrinne, "Media-independent Pre-authentication Supporting Secure Interdomain Handover Optimization," IEEE Wireless Communications, vol. 15, no. 2, 2008.

[47] A. Dutta, T. Zhang, Y. Ohba, K. Taniuchi, and H. Schulzrinne, "MPA Assisted Optimized Proactive Handoff Scheme," Proceedings of Mobile and Ubiquitous Systems: Networking and Services, 2005.

[48] T. Melia, F. Giust, R. Manfrin, A. de la Oliva, C. Bernardos, M. Wetterwald, "IEEE 802.21 and Proxy Mobile IPv6: A Network Controlled Mobility Solution," Proceedings of Future Network and Mobile Summit, 2011.

[49] W. Shi, B. Li, N. Li and C. Xia, "A Network Architecture for Load Balancing of Heterogeneous Wireless Network," Journal of Networks, vol. 6, no. 4, Apr. 2011.

[50] IEEE Std 802.21-2008, IEEE Standard for Local and Metropolitan Area Networks, Part 21: Media Independent Handover Services, January 2009.

[51] S. Buljore, H. Harada, S. Filin, P. Houze, K. Tsagkaris, O. Holland, K. Nolte, T. Farnham, and V. Ivanov, "Architecture and Enablers for Optimized Radio Resource usage in Heterogeneous Wireless Access Networks: The IEEE 1900.4 Working Group," IEEE Communications Magazine, January 2009.

[52] L. Gavrilovska, and V. Atanasovski, "Resource Management in Wireless Heterogeneous Networks (WHNs)," Telsiks, October 2009.

[53] Z. Sun and W. Wang, "Investigation of Co-operation Technologies in Heterogeneous Wireless Networks," J. Hindawi Publishing Corporation Computer Systems, Networks, and Communications, 2010.

[54] N. Dimitriou, L. Sarakis, D. Loukatos, G. Kormentzas, C. Skianis, "Vertical Handover Framework for Future Collaborative Wireless Networks," International Journal of Network Management, vol. 21, no. 6, pp. 548-564, Dec. 2011.

[55] E. Ong, J. Khan, K. Mahata, "Radio Resource Management of Composite Wireless Networks: Predictive and Reactive Approaches," IEEE Transactions on Mobile Computing, vol. 11, no. 5, May 2012.

[56] A. Oliva, C. Bernardos, M. Calderon, T. Melia, and J. Zuniga, "IP Flow Mobility: Smart Traffic Offload for Future Wireless Networks," IEEE Communications Magazine, vol. 49, no. 10, Oct 2011.

[57] C. Bernardos, "Proxy Mobile IPv6 Extensions to Support Flow Mobility," IETF, draft-ietf-netext-pmipv6-flow- mob-01, work in progress, Sep. 2011.

[58] 3rd Generation Partnership Project, "3GPP TS 23.402," Architecture Enhancements for Non-3GPP Accesses, 2011.

[59] W. Wolf et. al, "Multiprocessor System-on-Chip Technology," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 27, no. 10, pp. 1701-1713, Oct. 2008.

[60] A. Schranzhofer et. al, "Dynamic Power-Aware Mapping of Applications onto Heterogeneous MPSoC Platforms," IEEE Transactions on Industrial Informatics, vol. 6, no. 4, pp. 692-707, Nov. 2010.

[61] M. Shafique, L. Bauer, and J. Henkel, “REMiS: Run-time Energy Minimization Scheme in a Reconfigurable Processor with Dynamic Power-Gated Instruction Set,” in Computer-Aided Design – Digest of Technical Papers, pp. 55-62, Nov. 2009.

[62] Jean-Yves Le Boudec, ”Rate Adaptation, Congestion Control and Fairness: A Tutorial,” Ecole Polytechnique Federale de Lausanne(EPFL), Chapter 1, November 22, 2005.

[63] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, ”Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability” Journal of Operational Research Society, vol. 49, 1998.

[64] T. Girici, C. X. Zhu, J. R. Agre and A. Ephremides, “Proportional Fair Scheduling Algorithm in OFDMA-Based Wireless Systems with QoS Constraints,” Journal of Communications and Networks, vol. 12, no. 1, February 2010.

[65] M. Alasti, B. Neekzad, C. J. Hui, and R.Vannithamby, “Quality of Service in WiMAX and LTE networks,” IEEE Communications Magazine, vol. 48, no. 5, May 2010.

[66] I. Mansfield, “Smartphone Battery Life has Become a Significant Drain on Customer Satisfaction and Loyalty,” Available Online: http://www.cellular-news.com/story /53523.php, Last accessed on April 1, 2012.

[67] J. Liang, J. Chen, C. Liu, Y. Tseng and B. Lin, “On Tile-and-Energy Allocation in OFDMA Broadband Wireless Networks,” IEEE Communications Letters, vol. 15, no. 12, Dec. 2011.

[68] F. Meshkati, H. Poor, S. Schwartz, and R. Balan, “Energy-Efficient Resource Allocation in Wireless Networks with Quality-of-Service Constraints,” IEEE Transactions on Communications, vol. 57, no. 11, Nov. 2009.

[69] P. Serrano, M. Hollick, and A. Banchs, “On the Trade-Off between Throughput Maximization and Energy Consumption Minimization in IEEE 802.11 WLANs,” Journal of Communications and Networks, vol. 12, no. 2, April 2010.

[70] Y. Zhang, and K. Letaief, “Multiuser Adaptive Subcarrier-and-Bit Allocation with Adaptive Cell Selection for OFDM Systems,” IEEE Transactions on Wireless Communications, vol. 3, no. 4, September 2004.

[71] W. Rhee, and J. Cioffi, “Increase in Capacity of Multiuser OFDM System using Dynamic Subchannel Allocation,” Proceedings of IEEE Vehicular Technology Conference, May 2000.

[72] D. Tse, "Multiuser Diversity in Wireless Networks," Proceedings of Stanford Wireless Communications Seminar, April 2001.

[73] T. Saaty, "How to Make a Decision: The Analytic Hierarchy Process," European Journal of Operational Research, vol. 48, 1990.

[74] C. J. Chang, T. L. Tsai, and Y. H. Chen, "Utility and Game-Theory Based Network Selection Scheme in Heterogeneous Wireless Networks," Proceedings of IEEE Wireless Communications And Networking Conference, 2009.

[75] J. Antoniou and A. Pitsillides, "4G Converged Environment: Modeling Network Selection as a Game," Proceedings of 16th IST Mobile and Wireless Comm. Summit, 2007.

[76] D. Niyato and E. Hossain, "Dynamics of Network Selection in Heterogeneous Wireless Networks," Proceedings of IEEE Transactions on Vehicular Technology, vol. 58, no. 4, pp. 2008-2017, May 2009.

[77] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A Survey on Networking Games in Telecommunications," Computers and Operations Research, vol. 33, pp. 286–311, 2006.

[78] A. Rashwan, H. Elbadawy, H. Ali, "Comparative Assessments for Different WiMAX Scheduling Algorithms," Proceedings of World Congress on Engineering and Computer Science, October 2009.

[79] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance Anomaly of 802.11b," Proceedings of IEEE INFOCOM, 2003.

[80] E. Garcia, D. Viamonte, R. Vidal, and J. Paradells, "Achievable Bandwidth Estimation for Stations in Multi-Rate IEEE 802.11 WLAN Cells," Proceedings of IEEE Int'l Symp. World of Wireless, Mobile and Multimedia Networks, June 2007.

[81] L. Jiang, and S. Liew, "Proportional Fairness in Wireless LANs and Ad Hoc Networks," Proceedings of IEEE Wireless Communications and Networking Conference, 2005.

[82] A. Banchs, P. Serrano, and H. Oliver, "Proportional Fair Throughput Allocation in Multirate IEEE 802.11e Wireless LANs," Wireless Networks, vol. 23, no. 5, Oct. 2007.

[83] T. Joshi, A. Mukherjee, Y. Yoo, and D. P. Agrawal, "Airtime Fairness for IEEE 802.11 Multirate Networks," IEEE Transactions on Mobile Computing, vol. 7, no. 4, April 2008.

[84] TS 136 300 V9.3.0, LTE: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, March 2010.

[85] 3rd Generation Partnership Project, "TS 23.228 V11.3.0," IP Multimedia Subsystem (IMS); Stage 2 (Release 11), December 2011.

[86] J. Walrand, Economic Models of Communication Networks, Chapter 3 in Performance Modeling and Engineering. Z. Liu, C. Xia (Eds), Springer Publishing Company, 2008.

[87] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC Research Report TR-301, Sep. 1984.

[88] L.Shang, A.S.Kaviani, K.Bathala "Dynamic Power Consumption in Virtex II FPGA Family," ACM/SIDGA International Symposium on Field Programmable Gate Arrays, Feb. 2002.

[89] I. Kuon and J. Rose, "Measuring the Gap Between FPGAs and ASICs," in Proc. ACM/SIGDA 10th Int. Symp. Field-Programmable Gate Arrays, 2006.

[90] S. Lee, S. SungHoon, and N. Golmie, "An Efficient Power-Saving Mechanism for Integration of WLAN and Cellular Networks," IEEE Communication Letters, vol. 9, no. 12, Dec. 2005.

[91] S. Fowler, "Study on Power Saving Based on Radio Frame in LTE Wireless Communication System Using DRX," Joint Workshop of SCPA and SaCoNAS, 2011.

[92] J. Manweiler, R. Choudhury, "Avoiding the Rush Hours: WiFi Energy Management via Traffic Isolation," IEEE Transactions on Mobile Computing, vol. 11, no. 5, May 2012.

[93] L. M. Feeney, and M. Nilsson, "Investigating the Energy Consumption of Wireless Network Interface in an Ad Hoc Networking Environment," Proceedings of the IEEE Infocom conference, 2001.

[94] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications," Proceedings of the 9th ACM SIGCOMM Internet Measurement Conference, 2009.

[95] A. Saha, and A. Sinha, "An FPGA Based Architecture of a Novel Reconfigurable Radio Processor for Software Defined Radio," Proceedings of International Conference on Education Technology and Computer," 2009.

[96] J. Iyengar, P. Amer, and R. Stewart, "Concurrent Multipath Transfer using SCTP Multihoming over Independent End-to-End Paths," Transactions on Networking, vol. 14, no. 5, pp. 951-964, Oct. 2006.

[97] P. Viswanath, D. Tse, and R. Laroia, "Opportunitstic Beamforming using Dumb Antennas," IEEE Transactions on Information Theory, vol. 48, June 2002.

[98] J. Andrews, A. Ghosh, and R. Muhamed, Fundamentals of WiMAX: Understanding Broadband Wireless Networking, Prentice Hall, NJ, 2007.

[99] E. Forman, and K. Peniwati, "Aggregating Individual Judgments and Priorities with the Analytic Hierarchy Process," Elsevier European Journal of Operational Research, vol. 108, no. 1, 1998.

[100] R. Ramanathan, and L. Ganesh, "Group Preference Aggregation Methods Employed in AHP: An Evaluation and an Intrinsic Process for Deriving Members' Weightages," Elsevier European Journal of Operational Research, vol. 79, no. 2, 1994.

[101] J. D. Deaton, C. Wernz, and L. A. DaSilva, "Decision Analysis for Dynamic Spectrum Access Rules," Proceedings of IEEE Globecom, December 2011.

[102] R. Fourer, D. M. Gay, and B. W. Kernighan, AMPL: A Mathematical Programming Language. Brooks/Cole-Thomson Learning, 2003.

[103] CPLEX. ILOG CPLEX 10.0 User's Manual and Reference Manual. ILOG, S.A., 2006.

[104] T. Cover, and J. Thomas, "Elements of Information Theory," Wiley, New York, NY, 1991.

[105] Y. Bejerano, S. Han, and L. E. Li, "Fairness and Load Balancing in Wireless LANs using Association Control," in Proceedings of MOBICOM, pp. 315-329, 2004.

[106] W. Wang, X. Liu, J. Vicente, and P. Mohapatra, "Integration Gain of Heterogeneous WiFi/WiMAX Networks," IEEE Transactions on Mobile Computing, Dec. 2010.

[107] NIST Mobility Package, "Seamless and Security Project: Software Tools," Available Online: http://www.nist.gov/itl/antd/emntg/upload/MIH_module.pdf, Last accessed June 2012.

[108] T. Narten, E. Nordmark, W. Simpson, and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)," RFC 4861, IETF, Sep. 2007.

[109] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights," IEEE Transactions on Wireless Communications, vol. 12, no. 5, May 2013.